Two Geometric Inequalities in Harmonic Analysis

Stefán Ingi Valdimarsson

Doctor of Philosophy
University of Edinburgh
2006

for my family

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

(Stefán Ingi Valdimarsson)

Abstract

We study two geometric inequalities in harmonic analysis.

In the first part we study the Brascamp–Lieb inequality. We re-examine several of the approaches that have yielded results for this inequality and use them to derive new results. Specifically we prove an inequality involving the Hessian of the optimal transport map and use it to derive the generalised Brascamp–Lieb and reverse Brascamp–Lieb inequality with the methods of Barthe. Also, we extend the heat flow methods from Carlen, Lieb and Loss to give the form of all optimisers for the Brascamp–Lieb inequality and we use the induction on dimension method of Bennett, Carbery, Christ and Tao to prove a Brascamp–Lieb inequality for finite fields. Finally, we study the set of L^p-indices where the Brascamp–Lieb inequality holds and give alternative ways of describing it in several situations.

In the second part we study a multilinear analogue of fractional integration which has been studied in one form by Drury. We give the L^p bounds for it and find the optimal constant for this bound in the case with the most symmetries. We also determine all functions which are optimisers for this inequality. Finally, we study an analogue of this form which corresponds to the Hilbert transform. Here the finiteness of the form depends on cancellation properties in the kernel and we show how to define the form in terms of distributions. Then we prove L^p bounds for that form.

Acknowledgements

First and foremost, I would like to thank my supervisor, Tony Carbery, for his continuing support and encouragement at every step of creating this thesis and beyond. Also, thanks to other members of the analysis community at the University of Edinburgh during my time here for valuable discussions, in particular Jim Wright, Jon Bennett and my office-mates Sanjay Patel, Spyros Dendrinos, Neal Bez, Daniel Wilheim and Michael Kowalski.

Much of this thesis was written during a two-month stay at the University of Athens as an early stage researcher within the European network Phenomena in High Dimensions. Apostolos Giannopoulos deserves great thanks for going well beyond what could be expected of him in organising that stay and for many interesting discussions.

The financial support of the Overseas Research Students Awards Scheme is acknowledged.

Finally, I would like to thank my wife, Elínborg, for all her help, support and love during our time here in Edinburgh.

Table of Contents

Chapter 1 Introduction **2**

1.1 Overview 2

1.2 Rearrangements of functions 5

1.3 Mass transportation 9

1.4 The Brascamp–Lieb inequality 12

1.4.1 Conditions for finiteness of $C((B_j),(p_j))$ 14

Chapter 2 The Brascamp–Lieb polyhedron **17**

2.1 Introduction 17

2.2 The vertices of $\mathcal{S}$ 20

2.2.1 Mixed rank one and two 24

2.3 The facets of $\mathcal{S}$ 29

Chapter 3 The Brascamp–Lieb inequality on finite fields **33**

Chapter 4 Heat flow and optimisers for Brascamp–Lieb inequalities **39**

4.1 Monotonicity properties 39

4.2 Determination of optimisers 44

Chapter 5 The Hessian of the optimal transport potential **55**

5.1 Proof of Theorem 5.2 58

5.2 Proof of Theorem 5.4 65

Chapter 6 A multilinear generalisation of the Hilbert transform and fractional integration **68**

6.1 A singular integral 68

6.2 A fractional integral 82

Bibliography **93**

Chapter 1

Introduction

1.1 Overview

This thesis is about multilinear inequalities in harmonic analysis which have a geometric flavour.

The bulk of the thesis, Chapters 2 to 5, is devoted to the Brascamp–Lieb inequality

$$\int_H \prod_{j=1}^m f_j^{p_j}(B_j x)\,\mathrm{d}x \le C \prod_{j=1}^m \left(\int_{H_j} f_j \right)^{p_j} \tag{1.1}$$

where H and H_j are finite dimensional Hilbert spaces of dimensions n and n_j respectively, $B_j : H \to H_j$ are linear maps, p_j are non-negative numbers, C is a finite constant and f_j are non-negative functions. We shall refer to $((B_j), (p_j))$ as the Brascamp–Lieb datum for this inequality and (B_j) as the m-transformation for it.

In Chapter 2 we study the conditions that have been found to be necessary and sufficient for (1.1) to hold. If the maps B_j are fixed these conditions show that (1.1) holds if and only if the tuple (p_j) lies in a certain polyhedron $\mathcal{S}$. In Section 2.2 we search for the vertices of this polyhedron. We establish several general lemmas which aid this search and with those we list all vertices of $\mathcal{S}$ in the cases when the maps B_j have

1. common rank 1, see Theorem 2.1;

2. common rank $n-1$, see Theorem 2.2;

3. mixed rank 1 and 2, see Theorem 2.12; and

4. indicate how to treat the common rank $n-2$ case, see Remark 2.15.

One of the necessary conditions for (1.1) to hold is that

$$\dim V \le \sum_j p_j \dim(B_j V) \tag{1.2}$$

holds for all subspaces V of H. As there are an uncountable number of these spaces it is far from obvious how to arrive at a finite description of $\mathcal{S}$. We address this problem in Section 2.3 and show that it is sufficient to test (1.2) on spaces V in the lattice of $(\ker B_j)_{j=1}^m$. This is defined to be the smallest set which contains each $\ker B_j$ and is closed under intersection and vector-space addition. This lattice is countable and we give an algorithm which determines when all distinct conditions included in (1.2) have been found.

In Chapter 3 we study the analogue of (1.1) when the Hilbert spaces H and H_j are replaced by vector spaces over a finite field, $\mathbb{F}^n$ and $\mathbb{F}^{n_j}$. We prove that if this analogue holds with a constant C which is independent of the cardinality of the field $|\mathbb{F}|$ then C can be taken to be 1 and we give necessary and sufficient conditions for this to be the case. We refer to Theorem 3.2 for details. Furthermore, we use the relationship between the Brascamp–Lieb inequalities and convolutions to give a version of Lieb's theorem, Theorem 1.12 below, in the finite field case, see Theorem 3.3 for details.

In Chapter 4 we return to the Hilbert space problem and find all optimisers for (1.1). Our approach is based on arguments establishing the monotonicity of a certain heat flow related to inequality (1.1). This technique was introduced in the rank one case by Carlen, Lieb and Loss [16] and independently and in the general case by Bennett, Carbery, Christ and Tao [7]. When we look for the optimisers it is convenient to work with a Brascamp–Lieb datum that interacts nicely with the Hilbert space structure on H. This is the case for what is known as a geometric Brascamp–Lieb datum, following Ball [1], see Definition 1.13. In particular in this case, each H_j is a subspace of H and B_j is the orthogonal projection. We are able to see that the integrand on the left hand side of (1.1) factors into two terms, involving only B_jx and $B_j^{\perp}x$ respectively, where $B_j^{\perp}$ is the orthogonal projection onto $H_j^{\perp}$. The fact that this decomposition holds for each j yields a lot of structure which we exploit. See Theorem 4.5 for the optimisers in the case of a geometric datum and Theorem 4.8 for the general case which follows easily.

We note in particular that these theorems recover such different cases of equality as the multilinear Hölder's inequality, where each tuple of the form $(f_j) = (c_j f)$ with each c_j a constant and f an arbitrary integrable function gives an optimiser, and Young's inequality where the only optimisers are tuples of gaussians, $(f_j) = (e^{-\pi\langle \cdot, A_j \cdot\rangle})$ where the A_j's are certain positive definite linear transformations on H_j.

In Chapter 5 we revisit the method which Barthe has developed in [3] to study the Brascamp–Lieb inequalities and study how it applies to the generalised

Brascamp–Lieb inequalities as introduced in [7]. Barthe's proof also gives a reverse Brascamp–Lieb inequality and we derive a generalised reverse Brascamp–Lieb inequality. We discuss these results further in Section 1.3.

Finally, in Chapter 6 we study multilinear generalisations of the Hilbert transform and the Hardy–Littlewood–Sobolev theorem. Firstly we study the n-linear form

$$\Lambda(f_1, \dots, f_n) := \int_{(\mathbb{R}^{n-1})^n} \frac{f_1(x_1) \cdots f_n(x_n)}{K(x_1, \dots, x_n)} \, \mathrm{d}x_1 \, \mathrm{d}x_2 \dots \mathrm{d}x_n \tag{1.3}$$

with each $x_i \in \mathbb{R}^{n-1}$ and the kernel given by

$$K(x_1, \dots, x_n) = \det \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix}.$$

This multilinearises the Hilbert transform as for $n = 2$ we see that $\det \left(\begin{smallmatrix} 1 & 1 \\ x & y \end{smallmatrix} \right) = x - y$. As is the case with the Hilbert transform, the integral in (1.3) is not convergent in general but we address that issue in equation (6.3) and Lemma 6.1. We give a quite precise result for when

$$|\Lambda(f_1, \dots, f_n)| \leq C \|f_1\|_{L^{p_1}(\mathbb{R}^{n-1})} \cdots \|f_n\|_{L^{p_n}(\mathbb{R}^{n-1})} \tag{1.4}$$

holds in Theorem 6.2.

(Note that in this discussion, and in Section 1.2 below and Chapter 6 in general, we state our results in terms of L^p spaces and p_j in this theorem is the index of these spaces. In particular, $p_j \geq 1$ and we expect necessary and sufficient conditions to be linear constraints in $(\frac{1}{p_j})$. In the discussion on the Brascamp–Lieb inequality, on the other hand, we state our results in terms of L^1 functions, the powers p_j satisfy $p_j \leq 1$ and the necessary and sufficient conditions are linear constraints in (p_j).)

The analogue of the Hardy–Littlewood–Sobolev theorem comes from considering the integral in (1.3) with K replaced by K_α where

$$K_\alpha(x_1, \dots, x_n) = \left| \det \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \right|^\alpha$$

and $0 < \alpha < 1$. Drury [19] has proved that the resulting form, Λ_α, satisfies

$$\Lambda_\alpha(f_1, \dots, f_n) \leq C \|f_1\|_{p_\alpha} \dots \|f_n\|_{p_\alpha} \tag{1.5}$$

where $1/p_\alpha = 1 - \alpha/n$ using methods similar to the ones we use for Theorem 6.2. In Section 6.2 we use rearrangements of functions and the interplay between Λ_α and a closely related form on the sphere S^{n-1} to determine the best constant C in the estimate (1.5) and find all functions which give equality in the estimate with that constant.

In the remaining sections of this Introduction we will place these results in context and give the background results we will need from the literature. Specifically, in the next two sections we will introduce two highly useful techniques of passing between functions, namely rearrangements and mass transport. A third technique with the same spirit, heat flow, will be useful in Chapter 4. In the last section we discuss the portions of the theory of Brascamp–Lieb inequalities which will be relevant to us.

1.2 Rearrangements of functions

Let E be a subset of $\mathbb{R}^n$ of finite Lebesgue measure. We define the *symmetric rearrangement* of E to be the open ball in $\mathbb{R}^n$ centred at the origin which has the same measure as E and denote it by E^*. With this definition we can define the symmetric rearrangement of an integrable function f as

$$\mathcal{R}f(x) = f^*(x) = \int_0^\infty \chi_{\{|f(x)|>t\}^*}(x)\,\mathrm{d}t. \tag{1.6}$$

The measure of the set $\{x : |f(x)| > t\}$ in this definition is called the *distribution function* of f, denoted $\lambda(t)$. Since rearrangement, as we have defined it, leaves the measure invariant we see that f and f^* have the same distribution function. Therefore, any quantity which can be defined in terms of the distribution function is invariant under the rearrangement action. This holds in particular true for the L^p norm of f which may be defined as

$$\|f\|_p^p = -\int_0^\infty t^p\,\mathrm{d}\lambda(t).$$

We also note the representation

$$|f(x)| = \int_0^\infty \chi_{\{|f(x)|>t\}}(x)\,\mathrm{d}t. \tag{1.7}$$

of the modulus of f which compares directly to (1.6) and in many cases allows questions about the rearrangement of functions to be reduced to questions about rearrangement of sets. We can also see that this rearrangement preserves orderings of non-negative functions, in the sense that if $0 \leq f(x) \leq g(x)$ for all $x \in \mathbb{R}^n$ then also $f^*(x) \leq g^*(x)$.

We shall find it convenient to introduce a variant of this rearrangement which only rearranges in a single variable. Thus we define for a function f of n variables the *Steiner symmetrisation* with respect to the j-th coordinate direction as

$$\mathcal{R}_j f(x) = f^{*j}(x_1, \ldots, x_n) = \int_0^\infty \chi^*_{\{|f(x_1,\ldots,x_{j-1},\cdot,x_{j+1},\ldots,x_n)|>t\}}(x_j)\,\mathrm{d}t.$$

We see then that

$$\mathcal{R}_j f(x) = f^{*j}(x) = \int_0^\infty \chi^{*j}_{\{|f(x)|>t\}}(x)\,\mathrm{d}t$$

which again compares directly to (1.6). As before we have that $\|f\|_p = \|f^{*j}\|_p$ and the operator $\mathcal{R}_j$ is order preserving. We also note that f^{*j} decreases as the absolute value of the j-th coordinate increases.

There exist several rearrangement inequalities, that is, inequalities which relate an integral involving some non-negative functions to the same integral but with the rearrangement of the functions instead of the original functions. The simplest of these is the following.

Theorem 1.1. *Let f and g be non-negative integrable functions on $\mathbb{R}^n$. Then*

$$\int_{\mathbb{R}^n} f(x)g(x)\,\mathrm{d}x \le \int_{\mathbb{R}^n} f^*(x)g^*(x)\,\mathrm{d}x.$$

The proof is immediate, we use the representations (1.6) and (1.7) to see that we want to prove

$$\int_{\mathbb{R}^n}\int_0^\infty\int_0^\infty \chi_{\{f>t\}}(x)\chi_{\{g>s\}}(x)\,\mathrm{d}s\,\mathrm{d}t\,\mathrm{d}x \le \int_{\mathbb{R}^n}\int_0^\infty\int_0^\infty \chi^*_{\{f>t\}}(x)\chi^*_{\{g>s\}}(x)\,\mathrm{d}s\,\mathrm{d}t\,\mathrm{d}x.$$

By Fubini's theorem we see that it is enough to show that

$$|\{f>t\}\cap\{g>s\}| \le |\{f>t\}^*\cap\{g>s\}^*|$$

and this comes directly from the fact that

$$|\{f>t\}^*\cap\{g>s\}^*| = \min\{|\{f>t\}^*|, |\{g>s\}^*|\} = \min\{|\{f>t\}|, |\{g>s\}|\}$$

whereas in general we have

$$|\{f>t\}\cap\{g>s\}| \le \min\{|\{f>t\}|, |\{g>s\}|\}.$$

A more interesting theorem is the following Riesz's rearrangement inequality named after F. Riesz.

Theorem 1.2. *Let f, g and h be non-negative integrable functions on $\mathbb{R}^n$. Then*

$$I(f,g,h) \le I(f^*,g^*,h^*) \tag{1.8}$$

where

$$I(f,g,h) = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f(x)g(x-y)h(y)\,\mathrm{d}x\,\mathrm{d}y.$$

This may be extended to non-integrable functions in the sense that if the left hand side is infinite then so is the right hand side.

This theorem has been generalised to larger tuples of functions and more general linear maps by Brascamp, Lieb and Luttinger [9] as follows.

Theorem 1.3. *Let $f_1, \dots, f_m$ be non-negative functions on $\mathbb{R}^n$ and (b_{ij}) be a $k \times m$ matrix where $k \leq m$. Let*

$$I(f_1, \dots, f_m) = \int_{\mathbb{R}^n} \cdots \int_{\mathbb{R}^n} \prod_{j=1}^{m} f_j \left(\sum_{i=1}^{k} b_{ij} x_i \right) \mathrm{d}x_1 \dots \mathrm{d}x_k.$$

Then $I(f_1, \dots, f_m) \leq I(f_1^*, \dots, f_m^*)$.

The cases of equality have been studied, in particular by Lieb [24] and Burchard [12] and [13]. Lieb shows that if the g in Theorem 1.2 is strictly spherically decreasing then there is equality in (1.8) if and only if f and h are spherically decreasing functions, possibly translated by some common vector. Burchard studies several more general cases but when all of the functions f, g and h are completely unrestricted the condition she obtains for equality in (1.8) is difficult to check in any given case. It is much more tractable when the functions are characteristic functions.

An important case which lies between the ones of Riesz and Brascamp–Lieb–Luttinger is that of multiple convolutions where we let

$$\begin{aligned} I(f_1, \dots, f_{m+1}) &= \int_{\mathbb{R}^n} (f_1 * \cdots * f_m)(x) f_{m+1}(x) \, \mathrm{d}x \\ &= \int \cdots \int f_1(x_1) \dots f_m(x_m) f_{m+1} \left(\sum_{j=1}^{m} x_j \right) \mathrm{d}x_1 \dots \mathrm{d}x_m. \end{aligned} \tag{1.9}$$

Clearly this form falls under the scope of Theorem 1.3 but it had also been previously studied by Beckner [5], and for $m = 2$ and $n = 1$ this is Riesz's result. According to [5], Riesz remarks that it goes back to the methods of Hardy and Littlewood for rearrangement of series. The first proof on $\mathbb{R}^n$ seems to be due to Sobolev, see again [5].

Burchard [12] has found good conditions for the cases of equality in (1.9) when the functions are characteristic functions and also extended Lieb's result. We state that extension here with a slight added generality.

Lemma 1.4. *Assume that $f_1, \dots, f_{m+1}$ are non-negative functions on $\mathbb{R}^n$, f_{m+1} is symmetric decreasing and we have equality in the rearrangement inequality*

$$\begin{aligned} &\int_{(\mathbb{R}^n)^m} \prod_{j=1}^{m} f_j(x_j) f_{m+1}(b_1 x_1 + \cdots + b_m x_m) \, \mathrm{d}x_1 \dots \mathrm{d}x_m \\ &\qquad \leq \int_{(\mathbb{R}^n)^m} \prod_{j=1}^{m} f_j^*(x_j) f_{m+1}(b_1 x_1 + \cdots + b_m x_m) \, \mathrm{d}x_1 \dots \mathrm{d}x_m \end{aligned} \tag{1.10}$$

where $b_k \in \mathbb{R}$. Then there are vectors $a_1, \ldots, a_n \in \mathbb{R}^m$ such that $\sum b_i a_i = 0$ and $f_i(x_i) = f_i^*(x_i - a_i)$ for all $i = 1, \ldots, m$.

Burchard states her result with each $b_j = 1$ but by making the change of variables $x_j \mapsto b_j x_j$ the theorem reduces to that case.

Finally, let us note a different type of rearrangement inequality about how rearranging affects the difference of functions. At this level of generality it can be found in [27].

Theorem 1.5. *Let J be a non-negative convex function on $\mathbb{R}$ such that $J(0) = 0$ and let f and g be non-negative integrable functions on $\mathbb{R}^n$. Then*

$$\int_{\mathbb{R}^n} J(f^*(x) - g^*(x))\,\mathrm{d}x \leq \int_{\mathbb{R}^n} J(f(x) - g(x))\,\mathrm{d}x. \tag{1.11}$$

In particular we get that $\|f^* - g^*\|_{L^p} \leq \|f - g\|_{L^p}$ for $1 \leq p \leq \infty$.

Furthermore, if J is strictly convex and f satisfies $f = f^*$ then (1.11) *holds with equality if and only if $g = g^*$.*

Rearrangement inequalities have been invaluable in various problems relating to finding the best constant for various inequalities in harmonic analysis. We mention Beckner's discovery of the best constant in Young's inequality, [5]. Following the announcement of this result, Brascamp and Lieb gave a different proof [8], generalised the result to more than three functions and suggested further generalisations, namely what has become known as the Brascamp–Lieb inequality, see Section 1.4.

Another example is the Hardy–Littlewood–Sobolev inequality

$$\left|\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \frac{f(x)g(y)}{|x-y|^\alpha}\,\mathrm{d}x\,\mathrm{d}y\right| \leq C\|f\|_p\|g\|_q \tag{1.12}$$

which holds for $p, q > 1$ and $\alpha < 1$ such that $1/p + \alpha/n + 1/q = 2$. The best constant C for this equality has been found in the case $p = r$ by Lieb [25]. Later Carlen and Loss [17] revisited this problem using rearrangements and a certain type of conformal invariance and they were able to find all optimisers for (1.12), again in the case $p = r$.

In Section 6.2 where we consider a multilinear variant of (1.12) given by

$$\int_{(\mathbb{R}^{n-1})^n} \frac{f_1(x_1)\cdots f_n(x_n)}{\left|\det\left(\begin{smallmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{smallmatrix}\right)\right|^\alpha}\,\mathrm{d}x_1\,\mathrm{d}x_2\ldots\,\mathrm{d}x_n \leq C\|f_1\|_{p_\alpha}\cdots\|f_n\|_{p_\alpha} \tag{1.13}$$

with $1/p_\alpha = 1 - \alpha/n$ we are able to carry out a similar analysis to that of Carlen and Loss and find the best constant for this inequality and all the functions giving equality in the inequality with this constant.

1.3 Mass transportation

The problem of how to most efficiently transfer a mass of some shape, a pile of sand say, into another prescribed shape is a very natural one to consider and can be traced back to the paper of Monge [29] from 1781. His formulation can be stated as follows.

Problem 1.6. Let μ and ν be two probability measures on $\mathbb{R}^n$. Find a map $T : \mathbb{R}^n \to \mathbb{R}^n$ which transports μ to ν in the sense that

$$\int \zeta(T(y))\,\mathrm{d}\mu(y) = \int \zeta(y)\,\mathrm{d}\nu(y) \tag{1.14}$$

for all test functions ζ and which minimises the total distance travelled by the mass, that is the functional

$$\int c(x, T(x))\,\mathrm{d}x$$

where $c(x, y) = \|x - y\|$.

Although this may seem like a natural formulation of the problem there are several reasons why other forms might be considered better.

First of all, the problem as stated may not have a unique solution. Let us say for example in $\mathbb{R}$ that μ has half of its mass at -1 and the other half at 0 whereas ν has half of its mass at 0 and the other half at 1. Then the cost is the same whether we shift μ to the right by 1 to arrive at ν or if we leave the mass at 0 untouched and and move the mass at -1 to 1.

However if we take the cost function to be of the form $c(x, y) = J(\|x - y\|)$ where J is a strictly convex function then the transportation function is unique. On $\mathbb{R}$, moreover, it does not depend on the form of J. Informally we can say that the convex cost function dictates that in the optimal solution there is no rearrangement of mass, in the sense that if $x < y$ then $T(x) < T(y)$ and so the optimal transportation plan is to take the p-th percentile of the mass μ and place it in the p-th percentile of the mass ν.

There is still a question about the existence of solutions to Problem 1.6. For example, if we let μ be the point mass at 0 and ν have half its mass at -1 and the other half at 1 then there can be no function T which carries out the transportation. In practical terms, there is no question about how to get around this problem. We simply divide the pile of sand at 0 in two equal piles and ship one to the right and the other to the left.

This leads us to consider a more general transference plan as a probability measure π on the space $\mathbb{R}^n \times \mathbb{R}^n$ whose marginals are μ and ν respectively. This means that for any test function ζ

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} \zeta(x)\, \mathrm{d}\pi(x,y) = \int_{\mathbb{R}^n} \zeta(x)\, \mathrm{d}\mu(x) \tag{1.15}$$

and

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} \zeta(y)\, \mathrm{d}\pi(x,y) = \int_{\mathbb{R}^n} \zeta(y)\, \mathrm{d}\nu(y). \tag{1.16}$$

The transportation problem then becomes the following.

Problem 1.7 (Kantorovich's optimal transportation problem). Minimise

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} c(x,y)\, \mathrm{d}\pi(x,y)$$

subject to the conditions (1.15) and (1.16).

The problem of Monge is then the more restricted one of considering only those probability measures π which are supported in a graph of a function T.

Kantorovich studied the problem which bears his name in 1942 and in particular made the connection of viewing it in general as a continuous limit of a finite dimensional linear programming problem and moreover introduced a highly useful notion of duality to the problem. Incidentally, Kantorovich had also introduced the concept of linear programming a few years earlier and in the end received a Nobel prize in economics for his work.

We shall not be interested in the transport problem for general cost functions but shall rather specify the cost to be quadratic, that is $c(x,y) = \|x-y\|^2$. In this case the solution to Problem 1.7 has extremely useful geometric properties. Although these problems are often considered with very few regularity assumptions on μ and ν we will assume that these measures are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^n$. In this case it is known that there is a unique solution to the problem of Kantorovich which is at the same time a solution to Monge's problem. Furthermore, the mapping T can be written uniquely as $\nabla\phi$ where ϕ is a convex function. We shall refer to ϕ as the *optimal transport potential.*

A result very much in this direction goes back to Knott and Smith [23] and it was proved by Brenier [10] [11] under the additional assumption that μ and ν have finite second order moments by using the duality theory mentioned above. However, these authors also suggested that there might be a more direct way to prove these results and that programme was carried out by McCann [28] and he was thus able to eliminate the moment condition.

Caffarelli has provided a regularity theory for the potential ϕ. What will be relevant to us of these results is the following.

Theorem 1.8. *Let μ and ν be measures on $\mathbb{R}^n$ with density functions f and g respectively and assume that they belong to the class $C^{0,\alpha}(\mathbb{R}^n)$ of Hölder continuous functions with index α. Then the solution to Problem 1.7 with $c(x,y) = \|x-y\|^2$ is a measure π which is supported in the graph of the differential of a convex function ϕ and given explicitly by*

$$\mathrm{d}\pi(x,y) = \mathrm{d}\mu(x)\delta[y = \nabla\phi(x)]$$

where δ is the Dirac delta measure.

Furthermore, ϕ is of class $C^{2,\alpha}(\mathbb{R}^n)$.

In recent years the field of mass transportation has been extensively studied, both for itself and for the large number of application the methods have found. For a short introduction to this area we suggest the article of Ball [2] and for a longer one, the book of Villani [31].

Here we shall mention only one result which is due to Caffarelli, [14], [15].

Theorem 1.9. *Suppose the densities f and g have the form*

$$f = (\det B)^{-\frac{1}{2}} e^{-\pi\langle B^{-1}\cdot,\cdot\rangle} \qquad \text{and} \qquad g = Ce^{-\pi\langle B^{-1}\cdot,\cdot\rangle - H}$$

where B is a positive definite symmetric linear transformation, H is convex and C is chosen so that $\int g = 1$. Then the optimal transport potential ϕ satisfies

$$\mathrm{Hess}(\phi, x) \le I$$

where I is the identity transformation.

In Chapter 5 we shall generalise Theorem 1.9 as follows.

Theorem 1.10. *Suppose f and g are of the form*

$$f = (\det(B^{-1}G))^{\frac{1}{2}} e^{-\pi\langle B^{-1}G\cdot,\cdot\rangle} * \mu \qquad \text{and} \qquad g = Ce^{-\pi\langle B^{-1}A^{-1}\cdot,\cdot\rangle - H}$$

where A, G and B are positive definite symmetric linear transformations, μ is a probability measure on $\mathbb{R}^n$, H is convex and C is chosen so that $\int g = 1$. Suppose also that $AB \le GB = BG$. Then the optimal transport potential ϕ satisfies

$$\mathrm{Hess}(\phi, x) \le G.$$

The introduction of μ and G in this theorem makes it strictly stronger than the one of Caffarelli. This theorem makes the strategy of Barthe for studying the Brascamp–Lieb inequality stronger. To state the result, fix B to be the identity transformation and say that functions of the form which f has in Theorem 1.10 are of *class* G and that functions of the form which g has are of *inverse class* G. Then we have the following.

Theorem 1.11.

1. *(Generalised Brascamp–Lieb)*

$$\int_{\mathbb{R}^n} \prod_{j=1}^{m} f_j^{p_j}(B_j x)\,\mathrm{d}x \leq \frac{1}{\sqrt{D_{\mathbf{G}}}} \prod_{j=1}^{m} \left(\int f_j\right)^{p_j}$$

for all f_j *of class* G_j *and*

2. *(Generalised Reverse Brascamp–Lieb)*

$$\int_{\mathbb{R}^n}^{*} \sup\left\{\prod_{j=1}^{m} g_j^{p_j}(y_j) : \sum_{j=1}^{m} p_j B_j^* y_j = x, y_j \in \mathbb{R}^n\right\} \mathrm{d}x \geq \sqrt{D_{\mathbf{G}}} \prod_{j=1}^{m} \left(\int g_j\right)^{p_j}$$

for all g_j *of inverse class* G_j

where

$$D_{\mathbf{G}} = \inf_{A_j \leq G_j} \left\{ \frac{\det(\sum_{j=1}^{m} p_j B_j^* A_j B_j)}{\prod_{j=1}^{m} (\det A_j)^{p_j}} \right\}.$$

Here $\int^*$ denotes the outer integral as the supremum may not be measurable. We note that the first half of this theorem has already been seen in [7].

1.4 The Brascamp–Lieb inequality

The Brascamp–Lieb inequality (1.1) unifies and generalises several of the most central inequalities in analysis, among others the inequalities of Hölder, Young and Loomis–Whitney. It was first written down by Brascamp and Lieb in [8] where they give two questions as open problems. The first one is to find the necessary and sufficient conditions on the datum for (1.1) to hold and the second one is to determine when the best constant for (1.1) is attained by a tuple of gaussian functions, $f_j(x) = e^{-\langle x, A_j x\rangle}$ with each A_j a symmetric and positive semi-definite linear transformation.

Their motivation for asking this second question comes from [8] where they study special cases of inequality (1.1), the most general of these is to let $n_j = \tilde{n}$ for all j, assume that n is a multiple of $\tilde{n}$, that is that $n = \tilde{n}\tilde{m}$ for some integer $\tilde{m}$ and let $B_j : \mathbb{R}^n \to \mathbb{R}^{\tilde{m}}$ take $x = (x_1, \ldots, x_{\tilde{m}})$ to $(\langle a_1^j, x_1\rangle, \ldots, \langle a_{\tilde{m}}^j, x_{\tilde{m}}\rangle)$ where each a_k^j is a vector in $\mathbb{R}^{\tilde{n}}$. In these cases, they answer the second question positively so it became natural to ask whether it is true in the fully general case.

Let us say a few words about how they approach this problem because it has nice connections to the material in Section 1.2. The main argument is for the case when $\tilde{m} = 1$, the more general case follows by integrating in the variables $x_1, \ldots, x_{\tilde{m}}$ one after another. They are therefore interested in the integral

$$I = \int_{\mathbb{R}^n} \prod_{j=1}^{m} f_j^{p_j}(\langle a^j, x\rangle)\,\mathrm{d}x.$$

This they raise to the M-th power for some large integer M so they study the product

$$I^M = \int_{\mathbb{R}^{nM}} \prod_{j=1}^{m} \left(\prod_{k=1}^{M} f_j^{p_j}(\langle a^j, x_k\rangle) \right) \prod_{k=1}^{M} \mathrm{d}x_k.$$

They then use the general rearrangement inequality of Brascamp, Lieb and Luttinger to estimate this with

$$\int_{\mathbb{R}^{nM}} \prod_{j=1}^{m} F_j^{p_j}(\langle a^j, x_1\rangle, \ldots, \langle a^j, x_M\rangle) \prod_{k=1}^{M} \mathrm{d}x_k$$

where $F_j^{p_j}$ is the symmetrisation of $\prod_{k=1}^{M} f_j^{p_j}(\langle a^j, x_k\rangle)$.

In the case when each f_j is a simple function, in the sense that it takes only finitely many values, K, the symmetrisation F_j will be the positive combination of the characteristic function of finitely many balls centred at the origin. By passing the integral inside the large number of sums that this procedure has introduced they can reduce themselves to consider this last integral with a characteristic function of a ball instead of F_j. The connection with gaussian functions comes now when they estimate this integral with the characteristic function replaced by a suitable gaussian.

Then they have to estimate a sum of the norms of these gaussians by the right hand side of (1.1). Their estimate here is very crude; it involves K, the number of values each f_j takes. However, by letting the power M pass to infinity while keeping K fixed they are able to recover the desired estimate.

These matters were revisited some fifteen years later by Lieb in a paper titled *Gaussian kernels have only Gaussian maximizers*, [26]. In that paper Lieb considers operators between functions on $\mathbb{C}^n$ which are defined in terms of gaussian kernels, such as

$$(\mathcal{G}f)(x) = \int G(x,y)f(y)\,\mathrm{d}y$$

where

$$G(x,y) = \exp\left(- \left\langle \begin{pmatrix} x \\ y \end{pmatrix}, M \begin{pmatrix} x \\ y \end{pmatrix} + l \right\rangle \right)$$

with M a symmetric complex valued $2n \times 2n$ matrix such that $\Re M$ is positive semi-definite and l a vector in $\mathbb{C}^{2n}$.

As the title suggests, the main object of this paper is to show that if $\mathcal{G}$ is a bounded operator from $L^p(\mathbb{R}^n)$ to $L^q(\mathbb{R}^n)$ and some conditions on the values of p and q are satisfied then the only extremisers for $\mathcal{G}$ are gaussians.

The application of this theory, or rather a multilinear version of it, to the Brascamp–Lieb inequality is indirect and it does not show that gaussians are the only extremisers but it shows that gaussians exhaust the inequality in the following sense.

Theorem 1.12 (Lieb's Theorem). *Let $C((B_j),(p_j))$ be the smallest constant we can take in* (1.1) *so that it holds for all tuples (f_j) of integrable functions and let $C_g((B_j),(p_j))$ be the smallest constant we can take so that it holds for tuples of centred gaussians. Then*

$$C((B_j),(p_j)) = C_g((B_j),(p_j)). \tag{1.17}$$

1.4.1 Conditions for finiteness of $C((B_j),(p_j))$

Even with Lieb's Theorem, the crucial question still remains wide open, when is $C((B_j),(p_j))$ finite and what is its value? A practical case which was first studied by Ball and later generalised by Barthe is that of a geometric datum.

Definition 1.13. We say that a Brascamp–Lieb datum is *geometric* if $B_jB_j^* = \mathrm{Id}_{\mathbb{R}^n_j}$ for each j and

$$\sum_{j=1}^{m} p_jB_j^*B_j = \mathrm{Id}_{\mathbb{R}^n} \tag{1.18}$$

In this case it is shown by Ball [1] and Barthe [3] that (1.1) holds with $C = 1$ for all centred gaussians and that the tuple $(f_j) = (e^{-\langle\cdot,\cdot\rangle})$ is an extremiser which does attain that constant.

Barthe [3] also gave a necessary and sufficient conditions for (1.1) to hold in the case when each B_j has rank one and he determined all the functions such that (1.1) holds with equality. This was later re-examined by Carlen, Lieb and Loss [16] who introduced heat flow arguments to the theory of Brascamp–Lieb inequalities.

However, the general case was not settled until the two papers of Bennett, Carbery, Christ and Tao, [7] and [6]. They give necessary and sufficient conditions for (1.1) to hold. These are the following:

$$p_j \geq 0 \tag{1.19}$$

for all j,

$$\dim H = \sum_j p_j \dim H_j, \tag{1.20}$$

and

$$\dim V \leq \sum_j p_j \dim B_j V \tag{1.21}$$

for all V.

To see the second of these we test (1.1) on $f_{j,\lambda}(x) = \frac{1}{\lambda} f_j(\frac{x}{\lambda})$ and note that since the right hand side is independent of λ the same must hold true for the left hand side.

For the third one we take a subspace V of H and let f_j be the characteristic function of the coin shaped region which is the ϵ neighbourhood of the unit ball of $B_j V$ sitting in H_j. Then we have that $\int f_j = \epsilon^{\dim H_j - \dim B_j V}$ and that $\int \prod_j f_j^{p_j} \geq \epsilon^{\dim H - \dim V}$. Since ϵ can be arbitrarily small and (1.1) must hold independently of ϵ we get condition (1.21) from this with the aid of (1.20).

These conditions are proved to be sufficient by a relatively simple argument in [6] based on multilinear interpolation and induction on dimension. In retrospect it should not come as a surprise that interpolation plays a role in the Brascamp–Lieb inequality as it generalises among others Young's inequality and the classical derivation of that is based on interpolation. Also the positive part of Barthe's result in the rank one case can be recovered very easily with interpolation. The connection with gaussians only surfaced with the work of Beckner when he was looking for the best constant in Young's inequality. We give a strengthened version of the multilinear interpolation argument in Section 2.3.

In [7], among many other things, a structural theory for the Brascamp–Lieb inequality is developed. We will conclude this discussion by giving some details that will be relevant to us. In this discussion we shall assume that $p_j > 0$ for each j. This is a harmless assumption in the sense that any factor with power zero in (1.1) can be omitted from the inequality without affecting the value of the expressions on either side.

Definition 1.14. Let $B_j : H \to H_j$ and $B'_j : H' \to H'_j$ be linear transformations. We say that (B_j) and (B'_j) are *equivalent* if there exist invertible linear transformations $C : H' \to H$ and $C_j : H'_j \to H_j$ such that $B'_j = C_j^{-1} B_j C$.

By a simple change of variables we can see that if (f_j) is an extremiser for $((B_j), (p_j))$ then $(f'_j) = (f_j \circ C_j)$ is an extremiser for $((B'_j), (p_j))$ and it is clear that if each function in the tuple (f_j) is a gaussian then the same holds for each function in the tuple (f'_j). Also we have from [7] that if $((B_j), (p_j))$ is extremisable then it is gaussian extremisable. In the rank one case this goes back to [16]. Furthermore, any extremisable data is equivalent to a geometric data, see [7] again.

Definition 1.15.

1. We say that a subspace V of H is *critical* if V is neither $\{0\}$ nor H and (1.21) holds with equality for V.

2. We say that a pair of subspaces (V, W) of H is a *critical pair* if V and W are complementary in H, so $V + W = H$ and $V \cap W = \{0\}$, and $B_j V$ and $B_j W$ are complementary in H_j for each j.

We will conclude this introduction by stating a lemma extracted from [7] which clarifies the relationship between these concepts.

Lemma 1.16. *Let* $((B_j), (p_j))$ *be a Brascamp–Lieb datum such that* (1.20) *holds and* (1.21) *holds for any subspace* V *of* H. *Then we have the following:*

1. *Each component of a critical pair is a critical subspace.*

2. *If the datum is extremisable then for any critical subspace* V *of* H *there exists a complementary subspace* W *of* H *such that* (V, W) *is a critical pair.*

3. *If the datum is geometric then it is extremisable and we may take* W *to be the orthogonal complement of* V *in* H.

Chapter 2

The Brascamp–Lieb polyhedron

2.1 Introduction

In [7] it is shown that the Brascamp–Lieb inequality (1.1) holds for the datum $((B_j),(p_j))$ if and only if we have

$$\dim V \leq \sum_j p_j \dim(B_j V) \tag{2.1}$$

for all subspaces V of H, the scaling condition

$$\dim H = \sum_j p_j \dim(H_j) \tag{2.2}$$

holds and

$$p_j \geq 0 \tag{2.3}$$

for all j.

In this chapter we study the following problem. Let us fix the maps B_j. Then for which tuples (p_j) does the Brascamp–Lieb inequality hold, that is which tuples satisfy (2.1), (2.2) and (2.3)?

Since each of the conditions is a linear inequality or equality in the variables (p_j) and since the coefficients in (2.1) are dimensions of spaces which can only range through a finite set, it is clear that the set of tuples (p_j) such that these conditions hold is a convex set in $\mathbb{R}^m$ whose boundary consists of a finite number of hyperplanes. It is thus a polyhedron and we shall refer to it as the *Brascamp–Lieb polyhedron* for the m-transformation (B_j).

The scaling and positivity conditions (2.2) and (2.3) imply that this polyhedron lies in the intersection of a hyperplane and the first 2^m-tant in $\mathbb{R}^m$. What portion of this intersection the polyhedron occupies can vary greatly. In particular, for Hölder's inequality the conditions in (2.1) do not give any restrictions and the polyhedron is this whole intersection. On the other hand, (2.1) for the Loomis–Whitney inequality restricts the polyhedron to the one point set $(p_j)_{1\le j\le n} = (\frac{1}{n-1})_{1\le j\le n}$.

The conditions (2.1), (2.2) and (2.3) give a description of the Brascamp–Lieb polyhedron, $\mathcal{S}$, in the sense that if we want to check whether a particular point (p_j) belongs to $\mathcal{S}$ then we can do so by checking (p_j) against each one of these conditions and if it satisfies them all then the point belongs to the polyhedron. However, for two reasons it might be considered of benefit to give an alternative description. Firstly, the shape of the polyhedron can still seem quite unclear, in particular we do not have a result which says that the point (p_j) lies in the polyhedron if and only if it is of some prescribed form. Secondly, there is the question how many conditions are included in (2.1). Although, as we said above, it is only a finite number because the dimension of the spaces involved can only range through a finite set, it remains unclear how to get an exhaustive list of the conditions as it would seem to require examining each subspace V of H. We will address both of these problems in this chapter.

For the first problem, it is known by the Weyl–Minkowski theorem that a bounded polyhedron is a polytope, that is the convex hull of a finite set of points. Furthermore, it is a consequence of Carathéodory's theorem that each point in a bounded polyhedron can be written as a convex combination of the vertices of the polyhedron. Here we say that a point (q_j) is a vertex of a polyhedron if there exists a hyperplane such that the intersection of the hyperplane and $\mathcal{S}$ is the singleton $\{(q_j)\}$ and by writing (p_j) as a convex combination of the vertices we mean that (p_j) lies in the polyhedron if and only if we can write

$$p_j = \sum_{s=0}^{s_0} \lambda_s q_{s,j}$$

for all j, where $\lambda_s \ge 0$, $\sum_s \lambda_s = 1$ and q_s for $s = 1, \dots, s_0$ is an enumeration of the vertices. For these standard results in convexity see for example [4].

The problem of determining the vertices of $\mathcal{S}$ has until now only been resolved in the rank-one case. There we have the following result.

Theorem 2.1 (Rank one case, Barthe [3]). *Let $B_j x = \langle v_j, x\rangle$ for vectors v_j in H. Then (q_j) is a vertex of $\mathcal{S}$ if and only if $q_j = \chi_I(j)$ where χ_I denotes a characteristic function of an index set I such that $(v_j)_{j\in I}$ forms a basis for H.*

As noted in the Introduction, this result is reproved in [16] and [7].

In Section 2.2 we present a new analysis of the properties of the vertices which has the benefit that aside from yielding a new proof of the result of Barthe it makes it possible to determine the form of the vertices in several other cases.

Theorem 2.2 (Rank $n-1$ case). *Assume B_j all have rank $n-1$ and for each j let $\{v_j\}$ be a nonzero element in the kernel of B_j. Then (q_j) is a vertex of $\mathcal{S}$ if and only if $q_j = \frac{1}{n-1}\chi_I(j)$ where I is an index set such that $(v_j)_{j\in I}$ forms a basis for H.*

The main lemma for our treatment of these results is the following.

Lemma 2.3. *Let (q_j) be a vertex of $\mathcal{S}$. Then the support of q, $\{j|q_j \neq 0\}$, can have at most n elements where n is the dimension of H.*

Finally, we will also push the analysis further to give a description of the vertices in the case when each B_j has rank either 1 or 2.

In Section 2.3 we address the second problem mentioned above, how can we know which conditions are included in (2.1). To state the result we make the following definition.

Definition 2.4. Let $(V_k)_{k\in K}$ be a family of subspaces of a common space. Then the *lattice* of (V_k), denoted $\mathcal{L}_{(V_k)}$ is defined as follows

1. $V_k \in \mathcal{L}_{(V_k)}$ for each $k \in K$;

2. $V_1 \cap V_2, V_1 + V_2 \in \mathcal{L}_{(V_k)}$ for any $V_1, V_2 \in \mathcal{L}_{(V_k)}$.

We neither require $\{0\}$ nor the whole space to be elements of the lattice.

Remark 2.5. The lattice of a given family of spaces is the smallest set of spaces which contains each member of the family and is closed under the operations of set intersection and vector space addition; we say that the lattice is generated by the family.

Definition 2.6. For the m-transformation (B_j) we let $\mathcal{L}_{(B_j)}$ denote $\mathcal{L}_{(\ker(B_j))}$, the lattice generated by the kernels of B_j.

In Section 2.3 we prove the following theorem:

Theorem 2.7. *Let $((B_j),(p_j))$ be a Brascamp–Lieb datum. Then a necessary and sufficient condition for the the Brascamp–Lieb constant $C((B_j),(p_j))$ to be finite is that* (2.2) *and* (2.3) *hold and* (2.1) *holds for each subspace in $\mathcal{L}_{(B_j)}$.*

Remark 2.8. It is a comment of Michael Christ that by working through the induction proof of the Brascamp–Lieb inequality in [6] an algorithm which gives necessary and sufficient conditions for $C((B_j),(p_j))$ to be finite can be found. The proof that follows is along these lines and also establishes that the lattice $\mathcal{L}_{(B_j)}$ is sufficient. (The proof was done in close collaboration with Tony Carbery.)

However, even with Theorem 2.7 there remain some questions. Firstly, do we know that the number of elements in $\mathcal{L}_{(B_j)}$ is finite? The answer to this seems to be no in general, see [30] for an overview discussion on lattice theory, to which this question belongs. However, it is clear that the number of elements is countable and it is straightforward to generate a list of elements which we can check (2.1) on in sequence. So for computational purposes, a more important variant of this question is: how do we know when to stop, that is, when can we be sure that we have got a list of all the conditions included in (2.1)? We will address this question towards the end of Section 2.3.

2.2 The vertices of $\mathcal{S}$

Proof of Lemma 2.3. Assume U and W are two subspaces of H such that inequality (2.1) holds with equality for the point (q_j) of $\mathcal{S}$ and U and W. Then we get that

$$\begin{aligned}
&\sum_j q_j \dim(B_jU) + \sum_j q_j \dim(B_jW) \\
&\quad = \sum_j q_j \left(\dim(B_jU) + \dim(B_jW)\right) \\
&\quad = \sum_j q_j \left(\dim(B_jU \cap B_jW) + \dim(B_jU + B_jW)\right) \\
&\quad \geq \sum_j q_j \left(\dim(B_j(U \cap W)) + \dim(B_j(U + W))\right) \\
&\quad \geq (\dim(U \cap W) + \dim(U + W)) \\
&\quad = (\dim U + \dim W)
\end{aligned} \tag{2.4}$$

where we have used twice the fact that $\dim U + \dim W = \dim(U+W) + \dim(U \cap W)$ for any subspaces U and W. Also for the first inequality we have used that $\dim(B_jU + B_jW) = \dim(B_j(U+W))$ and $\dim(B_jU \cap B_jW) \geq \dim(B_j(U \cap W))$. The second inequality follows since (q_j) belongs to the polyhedron and therefore the condition (2.1) holds with (q_j) and both $U \cap W$ and $U + W$.

Since we are assuming that the beginning and end of this chain are equal, we must in fact have equality all the way. This tells us that we have equality in inequality (2.1) for $U \cap W$ and $U + W$ and that for all j such that $q_j > 0$ we have

$$\dim(B_jU) + \dim(B_jW) = \dim(B_j(U \cap W)) + \dim(B_j(U + W)). \tag{2.5}$$

We note that so far we have proved the following.

Lemma 2.9. *Let U and W be critical subspaces of H for a Brascamp–Lieb datum $((B_j), (p_j))$. Then $U \cap W$ and $U + W$ are also critical and for all j such that $p_j > 0$ we have that* (2.5) *holds.*

Now, if (q_j) is a vertex of $\mathcal{S}$ then we will have a set of indices, J, such that

$$q_j = 0 \qquad \text{for } j \notin J \tag{2.6}$$

and a collection of subspaces, $\mathcal{V}$, such that

$$\dim V = \sum_j q_j \dim(B_j V) \qquad \text{if } V \in \mathcal{V}. \tag{2.7}$$

A vertex of a polyhedron is the unique solution of the set of linear equations which the facets adjacent to the vertex satisfy. Thus, the system (2.6), (2.7) of linear equations determines the vertex (q_j) uniquely.

Let us now apply row operations to this system to simplify it. By subtracting the appropriate multiples of (2.6) from (2.7) we can substitute (2.7) with

$$\dim V = \sum_{j \in J} q_j \dim(B_j V) \qquad \text{for } V \in \mathcal{V}. \tag{2.8}$$

Now, take $U, W \in \mathcal{V}$. By the above discussion, we have $U \cap W, U + W \in \mathcal{V}$ as well and furthermore, the equality for W can be deduced from the equality for $U \cap W$, U and $U + W$ as follows.

$$\begin{array}{rcl}
\Big(& \dim(U \cap W) = \sum_{j \in J} q_j \dim(B_j(U \cap W)) & \Big) \\
+\Big(& \dim(U + W) = \sum_{j \in J} q_j \dim(B_j(U + W)) & \Big) \\
-\Big(& \dim U = \sum_{j \in J} q_j \dim(B_j U) & \Big) \\ \hline
=\Big(& \dim W = \sum_{j \in J} q_j \dim(B_j W) & \Big)
\end{array}$$

where we have used (2.5) to simplify the right hand side. This shows that we may remove the equation coming from W from (2.8) by row operations and thus without affecting the solution set.

Let us try and remove as many equations from (2.8) as we can. First of all, we may assume that $\{0\}$ is not in $\mathcal{V}$ as (2.8) is content free for that space. Let us then take a $U_1 \in \mathcal{V}$ such that no proper subspace of U_1 is in $\mathcal{V}$. Clearly such a space exists as we cannot have an infinite chain of nested subspaces in H. Define $\mathcal{V}_{U_1} := \{W \in \mathcal{V} : U_1 \subset W\}$. Then all the equalities for the subspaces in $\mathcal{V}$ can be deduced from the equalities for the subspaces in $\mathcal{V}_{U_1}$. To see this we note that if $W \in \mathcal{V} \setminus \mathcal{V}_{U_1}$ then $W \cap U_1 = \{0\}$ so the equality for W can be deduced from the equalities for U_1 and $U_1 + W$ which are elements of $\mathcal{V}_{U_1}$.

Next, let $U_2 \in \mathcal{V}_{U_1}$, $U_2 \neq U_1$ be such that no subspace $W \in \mathcal{V}_{U_1}$ lies properly between U_1 and U_2. Then as in the last paragraph we see that all equalities for subspaces in $\mathcal{V}_{U_1}$ can be deduced from the equalities for the subspaces in $\mathcal{V}_{U_2}$ and the equality for U_1. Continuing this process, we get a flag $U_1 \subsetneq U_2 \subsetneq \cdots \subsetneq U_s$ such that all the equalities for the subspaces in $\mathcal{V}$ can be deduced from the equalities for the spaces in this chain.

Thus we have seen that by using row operations we can remove all the equations from (2.8) except the ones coming from this flag, which we shall refer to as $\mathcal{U}$, and still have left the linear system

$$q_j = 0 \qquad \text{for } j \notin J; \tag{2.9}$$

$$\dim U = \sum_j q_j \dim(B_j U) \qquad \text{if } U \in \mathcal{U} \tag{2.10}$$

which is equivalent to the original one. Since H is n-dimensional, $\mathcal{U}$ can have at most n elements so the number of equations in (2.10) is at most n. However, since the system (2.9), (2.10) is a linear system which has a unique solution in $\mathbb{R}^m$, there must be at least m equations in the system. Therefore, there must be at least $m - n$ elements not in the set J and so the solution to the system (q_j) can have at most n non-zero elements.

This completes the proof of the lemma. □

The next lemma partly addresses the question how one can check that a particular point is a vertex.

Lemma 2.10. *Let a Brascamp–Lieb datum $((B_j), (p_j))$ be given and assume that $\mathcal{U} = (U_1, \ldots, U_s)$ is a flag in H, that is $U_1 \subsetneq U_2 \subsetneq \cdots \subsetneq U_s = H$, such that* (2.10) *holds. Assume also that the inequality* (2.1) *holds for any space $\tilde{W}$ which can be added into the flag.*

Then inequality (2.1) *holds for any subspace W of H so the Brascamp–Lieb inequality holds for this datum.*

Remark 2.11. If $\mathcal{U}$ is a maximal flag we cannot add any subspace to the flag so if we have a vector (q_j) for which (2.10) holds for a maximal flag $\mathcal{U}$ then (q_j) is a vertex of the Brascamp–Lieb polyhedron. This is however not a necessary condition for (q_j) to be a vertex, see Remark 2.14.

Proof of Lemma 2.10. If we re-examine the calculations in (2.4) we see that if we assume that (2.1) holds for $U \cap W$ and $U + W$ and it holds with equality for U then we get that (2.1) holds for W.

Let us now define $t_0 \in \{0, \ldots, s\}$ such that $U_{t_0} \subset W$ but $U_{t_0+1} \not\subset W$. To ensure that t_0 is well-defined we allow it to take the value 0 in which case we define $U_0 = \{0\}$. We see that if (2.1) holds for $W \cap U_{t_0+1}$ and $W + U_{t_0+1}$ then it holds for W. Since $U_{t_0} \subset W \cap U_{t_0+1} \subset U_{t_0+1}$ we see that (2.1) holds for $W \cap U_{t_0+1}$ by assumption. For $W + U_{t_0+1}$ we argue inductively. We note that $W + U_{t_0+1} \supset U_{t_0+1}$ so we can repeat this process for that space, that is find a $t_1 > t_0$ such that $U_{t_1} \subset W + U_{t_0+1}$ but $U_{t_1+1} \not\subset W + U_{t_0+1}$ and then (2.1) for $W + U_{t_0+1}$ will follow from the condition for $(W + U_{t_0+1}) \cap U_{t_1+1}$ which lies between U_{t_1} and U_{t_1+1} and the condition for $W + U_{t_1+1}$. This process will give us a flag $U_{t_0} \subset \cdots \subset U_{t_r}$ which is a subflag of the flag $\mathcal{U}$ and can therefore not contain more than s elements. Furthermore, this flag has the property that to confirm that (2.1) holds for W we need only to check that (2.1) holds for spaces V such that $U_t \subset V \subset U_{t+1}$ with $t \in \{t_0, \ldots, t_r\}$. Since W was arbitrary we have proved the lemma. □

Let us now list all the possible vertices in several cases. First let us assume that all the maps B_j have the same rank.

Proof of Theorem 2.1. As before, we let (q_j) be a vertex of the polyhedron and J be the set of indices j such that $q_j > 0$. If v_j for $j \in J$ do not span H then we do not have a solution to the system (2.1), (2.2) and (2.3). To see this, let V be a subspace of codimension 1 which contains v_j for all $j \in J$. Then $V^\perp$ lies in the kernel of all the relevant B_j. Therefore, testing (2.1) on $V^\perp$ gives $1 = \dim V^\perp \le \sum_j q_j \dim(B_j V^\perp) = 0$ which is impossible.

This and Lemma 2.3 shows that $|J| = n$ and for each $j \in J$ there is a vector v_j such that $v_j \in \ker B_{j'}$ for all $j' \in J \setminus \{j\}$ but $v_j \notin \ker B_j$. Define

$$U_j = \sum_{\substack{j' \in J \\ j' \le j}} \operatorname{span}(v_{j'}).$$

Then $\mathcal{U} = (U_j)_{j \in J}$ is a maximal flag in H. With these definitions of J and $\mathcal{U}$ we can see that (2.9) and (2.10) have the unique solution $q_j = 1$ for $j \in J$ and $q_j = 0$ otherwise. The note following Lemma 2.10 therefore gives that each vector of this form is a vertex of the polyhedron. □

Proof of Theorem 2.2. With (q_j) and J as before, we first note that if $\ker B_j$ for $j \in J$ do not span H then we do not have a solution to the system (2.1), (2.2) and (2.3) as can be seen from testing (2.1) on a space V such that $\sum_{j \in J} \ker B_j \subset V$ and $\dim V = n - 1$. This gives $n - 1 = \dim V \le \sum_j q_j \dim(B_j V) = (n-2) \sum_j q_j$ whereas the scaling condition (2.2) gives $n = \sum_j q_j (n-1)$.

From this and Lemma 2.3 we then see that $|J| = n$. Also, if we define

$$U_j = \sum_{\substack{j' \in J \\ j' \le j}} \ker B_{j'}$$

then $\mathcal{U} := (U_j)_{j\in J}$ is a maximal flag in H. The set of equations (2.10) for this flag becomes

$$s_j = \sum_{\substack{j'\in J\\ j'\le j}} q_{j'}(s_j-1) + \sum_{\substack{j'\in J\\ j'>j}} q_{j'}s_j \qquad j\in J$$

where $s_j := |\{j'\in J | j'\le j\}|$. Since the number of terms in the first sum is s_j and the number of terms in the last sum is $n-s_j$ it is evident that $q_j = \frac{1}{n-1}$ for $j\in J$ is a solution. Since the system has rank n this is the only solution and since the flag is maximal we get a vertex for the polyhedron. □

2.2.1 Mixed rank one and two

We can push this analysis further and examine the mixed rank case when each B_j has rank 1 or 2. Again, we assume (q_j) is a vertex of $\mathcal{S}$ and J and $\mathcal{U} = (U_1 \subsetneqq U_2 \subsetneqq \cdots \subsetneqq U_s)$ are such that (2.9) and (2.10) hold.

By subtracting the equation for U_{k-1} from the equation for U_k we see that we can replace (2.10) with

$$\dim(U_k/U_{k-1}) = \sum_j q_j(\dim(B_jU_k) - \dim(B_jU_{k-1})) \tag{2.11}$$

for $U_k\in\mathcal{U}$, $k\ge 1$ and with $U_0=\{0\}$. In this set of equations we note that the coefficients multiplying q_j sum up to the rank of B_j and the constant coefficients sum up to $\dim H$. Therefore, if we let J_1 and J_2 be the set of indices from J for the rank 1 and 2 transformations in the set $\{B_j | j\in J\}$ respectively and let m_1 and m_2 be the number of elements in these sets then the sum of the elements in the coefficient matrix of (2.11) equals m_1+2m_2. Furthermore, since the set of equations (2.11) uniquely determines $(q_j)_{j\in J}$ and $|J| = m_1+m_2$ we get that $s\ge m_1+m_2$.

Now, for each $j\in J_1$ the coefficients of q_j in (2.11) must all be 0 except one which must be 1. Therefore, at most m_1 of the equations can contain a non-zero coefficient for an element q_j with $j\in J_1$ and these equalities must contain at least m_1 of the non-zero coefficients in the matrix.

There are now two cases, either there is equality in each step of this calculation, that is there are exactly m_1 of the equations which have a non-zero coefficient for an element q_j with $j\in J_1$ and these equations have only these non-zero coefficients and there are exactly m_2 equations left which have all of the non-zero coefficients for the q_j with $j\in J_2$ which sum up to $2m_2$. Moreover, each of these m_2 equations must have either one coefficient equal to 2 and all other 0 or two coefficients equal to 1 and all other 0. Otherwise, if any of this does not hold, then, by the pigeonhole principle, there must be an equation among these, all of whose coefficients are zero except one, for q_j with $j \in J_2$, which must be one.

Note that we have made heavy use of the fact that the coefficients in (2.11) must all be non-negative integers and we may assume that no equation has zero coefficients in front of all the q_j as such an equation can be removed from the linear system and the corresponding subspace can be removed from the flag.

In the first case, we get that for each $j \in J_1$, the relevant equation from (2.11) takes the form $1 = q_j$. The left hand side must be 1 as we know that $0 < q_j \leq 1$ for each $j \in J$. Let us say that this is the equation coming from the quotient U_{k_j}/U_{k_j-1}. From the fact that $\dim(B_{j'}U_{k_j}) = \dim(B_{j'}U_{k_j-1})$ for all $j' \neq j$ we get that the intersection of $\ker B_{j'}$ with $U_{k_j} \setminus U_{k_j-1}$ is non-empty. Now, $U_{k_j-1} \subset \ker B_j$ whereas $U_{k_j} \setminus U_{k_j-1}$ contains no vectors in $\ker B_j$ so we see that $\ker B_{j'} \cap (H \setminus \ker B_j)$ is non-empty for any $j' \neq j$. Since $\dim \ker B_j = n-1$ we get by testing (2.1) on $\ker B_j$ that

$$\dim(H) - 1 \leq \sum_{j' \neq j} q_{j'} \dim(B_{j'}H). \tag{2.12}$$

Since we know that we have equality in (2.1) for H and since we have $q_j = 1$ we get by subtraction that (2.12) must in fact be an equality.

All in all, we get by repeating this process, rearranging and carrying out the reductions in the proof of Lemma 2.3 again that there exists a subspace H_1 of H such that $\dim B_j H_1 = 2$ for all $j \in J_2$ but H_1 lies in the kernel of all B_j for $j \in J_1$. Furthermore, H_1 is $n - m_1$ dimensional, the cosets $v_j + H_1$, $v_j \in \operatorname{im} B_j$ for $j \in J_1$, form a basis for H/H_1 and $q_j = 1$ for all $j \in J_1$.

So we are left with a flag in H_1 and m_2 equations associated with it, all of whose non-zero coefficients are for q_j with $j \in J_2$. If we have that one of these equations has only one non-zero coefficient, which must then be 2, then that equation must take the form $2 = 2q_j$. This we see since the left hand side cannot be larger than 2 as q_j is at most 1 and since we must always have

$$\dim(U_k/U_{k-1}) \geq \dim(B_j U_k) - \dim(B_j U_{k-1})$$

so the coefficient on the left hand side must be as large as any coefficient on the right hand side. Now, in the same way as before with the rank one spaces we get that there exists a subspace H_2 of H_1 and a flag in H_2 such that all of the equations associated to this flag have the form

$$t_{j,j'} = q_j + q_{j'} \tag{2.13}$$

for some $j, j' \in J_{2,1} \subset J_2$ and $t_{j,j'} \in \{1, 2\}$. Then if we let $J_{2,2} := J_2 \setminus J_{2,1}$ and for each $j \in J_{2,2}$ we let $\{v_{j,1}, v_{j,2}\}$ be a basis for $\operatorname{im} B_j$ then the set of cosets

$\{v_{j,l}+H_2 | j \in J_{2,2}$ and $l = 1,2\}$ forms a basis for H_1/H_2 and $q_j = 1$ for all $j \in J_{2,2}$. We also note that the flag we get by adding the span of the vectors from this basis one by one to the subspace H_2 is a maximal flag between H_2 and H_1 and we have equality in (2.11) for each step.

Now, if we have an equation in the set (2.13) with $t_{j,j'} = 2$ then we must have $q_j = q_{j'} = 1$ as neither can be greater than 1. Then we can insert a space into the flag which splits the single equation into the two equations and those equations are of the form we originally split off from the main argument. We will deal with these equations in the next paragraph but one and see that the index set of those should properly be considered as part of $J_{2,2}$.

When that rearrangement has been done we can thus get that all the equations concerning q_j with $j \in J_{2,1}$ have the form $1 = q_j + q_{j'}$. Let us define a relation on $J_{2,1}$ such that j is related to j' if there is an equation of the form $1 = q_j + q_{j'}$ with these j, j'. If we draw the graph of this relation then each vertex j will have exactly two edges connected to it. Therefore we can see that the graph will be a collection of disjoint circles. Let us examine one of these circles. We can write all of the equations relating to the vertices in this circle in the form

$$\begin{aligned}
q_{j_1} + q_{j_2} \qquad\qquad\qquad\qquad &= 1 \\
q_{j_2} + q_{j_3} \qquad\qquad\quad &= 1 \\
q_{j_{l-1}} + q_{j_l} &= 1 \\
q_{j_1} \qquad\qquad\qquad\qquad + q_{j_l} &= 1.
\end{aligned} \tag{2.14}$$

The number of equations in this list is the same as the number of variables. However, if there is an even number of equations then the sum we get by adding the even numbered equations is the same as the sum we get by adding the odd numbered equations and so this system does not have a unique solution, contrary to our assumptions. Therefore, the number of equations in each circle is odd and in that case the system has a unique solution, which clearly is $q_j = \frac{1}{2}$ for all $j \in J_{2,1}$. We note that as the left hand side of these equations is always 1, the flag they come from must be maximal.

Finally, let us look at the other case, where one of the equations in (2.11) is of the form $1 = q_j$ with $j \in J_2$. Since the sum of the coefficients in front of q_j equals 2 there must be another equation with the term q_j. Either, it also takes the form $1 = q_j$ or the form $t = q_j + Q$ where $t > 1$ is an integer and Q stands for terms with $q_{j'}$, $j' \in J_2 \setminus \{j\}$. Let us first examine the second case. Assume that it comes from (2.11) with U_{k_j}/U_{k_j-1} where the codimension of U_{k_j-1} in U_{k_j} is t. Since the coefficient multiplying q_j is 1 we get that there are $t-1$ independent vectors in the intersection of $\ker B_j/U_{k_j-1}$ and U_{k_j}/U_{k_j-1}. Let $\tilde{U}$ denote the vector sum of the span of these and U_{k_j-1}. By testing (2.1) on $\tilde{U}$ and subtracting (2.1) on U_{k_j-1} which we know gives an equality we get that $t-1 \leq Q'$ where Q' denotes the contribution to this sum from terms $q_{j'}$, $j' \in J_2 \setminus \{j\}$. Now we get the chain of inequalities

$$t = 1 + (t-1) \leq q_j + Q' \leq q_j + Q = t$$

and so we must have equality all the way and in particular this shows that we may add $\tilde{U}$ to the flag which gives equalities and assume that both equalities involving q_j take the form $1 = q_j$.

For the purpose of determining the vector (q_j) uniquely these two identical equations do the same as the single equation $2 = 2q_j$. We can therefore merge them and remove one space U_k from the flag $\mathcal{U}$. This shows that we may assume that the only equation involving this q_j has the form $2 = 2q_j$ and this we had already analysed above.

All in all we have proved the following.

Theorem 2.12 (Mixed rank 1 and 2). *Let B_j for $j \in J_1$ be rank 1 linear transformations from H and let B_j for $j \in J_2$ be rank 2 linear transformations. Then (q_j) is a vertex of S if and only if the following holds*

1. *$q_j = 1$ for all $j \in J_1$;*
2. *the set J_2 can be divided into two sets $J_{2,1}$ and $J_{2,2}$ such that*
   - *$q_j = \frac{1}{2}$ for all $j \in J_{2,1}$ and*
   - *$q_j = 1$ for all $j \in J_{2,2}$ and*
3. *the indices in $J_{2,1}$ can be split into classes such that the equations for each class take the form (2.14) and the number of indices in each class is odd.*
4. *There exists a maximal flag $U_1 \subsetneq \cdots \subsetneq U_n$ in H and numbers $0 \leq s_1 \leq s_2 \leq n$ such that*
   - *$\dim B_j U_{s_1} = 2$ for all $j \in J_{2,1}$ but $U_{s_1} \subset \ker B_j$ for all $j \in J_{2,2}$ and $j \in J_1$; and*
   - *$\dim B_j U_{s_2} = 2$ for all $j \in J_{2,2}$ but $U_{s_2} \subset \ker B_j$ for all $j \in J_1$.*

Remark 2.13. From the proof of the theorem it is clear that we may rearrange the flag so that the equations for q_j with $j \in J_{2,2} \cap J_1$ come in any order. However, this is not the case for U_{s_1}. In fact there might be only one way of choosing this maximal flag for U_{s_1}. An example of such a configuration is with $\dim H = 5$ and B_j for $j = 1, \ldots, 5$ are the rank two projections onto $\langle e_1, e_2 + e_3 \rangle$, $\langle e_1, e_4 \rangle$,

$\langle e_2 + e_1, e_4 + e_3\rangle$, $\langle e_2, e_5\rangle$ and $\langle e_3, e_5 + e_4\rangle$ respectively, where $\{e_i\}_{i=1,\dots,5}$ is a orthonormal basis for H and the angled brackets denote the span of the relevant vectors. Then the only maximal flag for which we have equality is

$$\langle e_5\rangle \subset \langle e_4, e_5\rangle \subset \langle e_3, e_4, e_5\rangle \subset \langle e_2, e_3, e_4, e_5\rangle \subset \langle e_1, e_2, e_3, e_4, e_5\rangle.$$

Remark 2.14. In the cases we have looked at, all of the vertices have had associated with them flags of maximal length. However, this is not the case in general as can be seen from the following example. We take H of dimension 8 with an orthonormal basis $(e_i)_{i=1,\dots,8}$. For $j = 1, \dots 4$ we take B_j to be the orthogonal projections onto the spaces $\langle e_1, e_2, e_5\rangle$, $\langle e_2, e_4, e_7\rangle$, $\langle e_1 + e_2, e_6, e_8\rangle$ and $\langle e_3 + e_4, e_5 + e_6, e_7 + e_8\rangle$ respectively. Then we have the flag

$$\langle e_1, e_2\rangle \subset \langle e_1, e_2, e_3, e_4\rangle \subset \langle e_1, e_2, e_3, e_4, e_5, e_6\rangle \subset \langle e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8\rangle$$

for which (2.11) becomes

$$\begin{aligned} p_1 + p_2 + p_3 &= 2 \\ p_1 + p_2 + p_4 &= 2 \\ p_1 + p_3 + p_4 &= 2 \\ p_2 + p_3 + p_4 &= 2 \end{aligned}$$

which has the solution $p_1 = p_2 = p_3 = p_4 = \frac{2}{3}$. It is straightforward to confirm that the inequality (2.1) is satisfied for any subspace V of H as from Lemma 2.10 we know that we need only to check it for subspaces which can be placed into the flag. However, no linear combination of the p_j with non-negative integer coefficients can equal 1 so there can be no one-dimensional subspace of H which has equality in (2.1).

Remark 2.15. If all the maps B_j have rank k then (2.2) gives that

$$\sum_j p_j = n/k \tag{2.15}$$

and we can rewrite (2.1) as

$$\dim V \le \sum_j p_j \dim(B_j V) = \sum_j p_j(\dim V - \dim(\ker B_j \cap V))$$

which says

$$\sum_j p_j \dim(\ker B_j \cap V) \le \frac{n-k}{k} \dim V. \tag{2.16}$$

We can carry out the analysis of this section with the conditions (2.3), (2.15) and (2.16) and in particular we can recover a theorem similar to Theorem 2.12 for the case when all B_j have rank $n-2$.

2.3 The facets of $\mathcal{S}$

We begin this section with a proof of Theorem 2.7.

Proof. The necessity of the conditions follows immediately from [7] as they are a subset of the necessary conditions established there.

To show that the conditions are sufficient we use induction on $n+m$, where $n = \dim H$ and m is the degree of multilinearity of the form. For the base case we consider $m = 1$. Then testing (2.1) on $\ker B_1$ gives that $\dim \ker B_1 = 0$ so B_1 is surjective and then the scaling condition gives $\dim H_1 = \dim H$ and $p_1 = 1$. We see then the inequality evidently holds with equality if we take $C(B_1, p_1) = (\det B_1)^{-1}$.

For the inductive step we take a datum $((B_j), (p_j))$ and assume that the result holds for each datum for which the quantity $m+n$ is smaller.

As before, the conditions (2.2), (2.3) along with (2.1) for $V \in \mathcal{L}_{(B_j)}$ define a bounded convex polyhedron in $\mathbb{R}^m$ and by multilinear interpolation, to show that the result holds everywhere in this polyhedron is is enough to establish it at each vertex of it. As we have already dealt with the case $m = 1$ we may assume $m > 2$ and then we get that at a vertex more than one of the linear inequalities defining the polyhedron must be satisfied with equality.

There are now two cases. Either we have $p_{j_0} = 0$ for some j_0 or there is a space $U \in \mathcal{L}_{(B_j)} \setminus \{\{0\}, H\}$ such that

$$\dim U = \sum_j p_j \dim(B_j U). \tag{2.17}$$

In the first case we see that we may write the Brascamp–Lieb inequality without referring to j_0 and the result thus follows from the induction hypothesis since the degree of multilinearity has been reduced.

In the second case we can factor the Brascamp–Lieb form in the following way: Define

$$\begin{aligned}
\tilde{B}_j &: U \to B_j U : x \mapsto B_j x \\
\tilde{\tilde{B}}_j &: U^\perp \to (B_j U)^\perp : x \mapsto \Pi_{(B_j U)^\perp} B_j x \\
\Gamma_j &: U^\perp \to B_j U : x \mapsto \Pi_{B_j U} B_j x
\end{aligned}$$

where $\Pi_{(B_j U)^\perp}$ and $\Pi_{B_j U}$ denote the orthogonal projections onto the relevant spaces. Then we can calculate

$$
\begin{aligned}
\int_H \prod_{j=1}^m f_j^{p_j}(B_j x)\,\mathrm{d}x &= \int_{U^\perp}\int_U \prod_{j=1}^m f_j^{p_j}(\tilde{B}_j\tilde{x} + B_j\tilde{\tilde{x}})\,\mathrm{d}\tilde{x}\,\mathrm{d}\tilde{\tilde{x}} \\
&\le C((\tilde{B}_j),(p_j)) \int_{U^\perp} \prod_{j=1}^m \left(\int_{B_jU} f_j(\tilde{y} + B_j\tilde{\tilde{x}})\,\mathrm{d}\tilde{y} \right)^{p_j} \mathrm{d}\tilde{\tilde{x}} \\
&= C((\tilde{B}_j),(p_j)) \int_{U^\perp} \prod_{j=1}^m \left(\int_{B_jU} f_j(\tilde{y} + \Gamma_j\tilde{\tilde{x}} + \tilde{\tilde{B}}_j\tilde{\tilde{x}})\,\mathrm{d}\tilde{y} \right)^{p_j} \mathrm{d}\tilde{\tilde{x}} \\
&= C((\tilde{B}_j),(p_j)) \int_{U^\perp} \prod_{j=1}^m \left(\int_{B_jU} f_j(\tilde{y} + \tilde{\tilde{B}}_j\tilde{\tilde{x}})\,\mathrm{d}\tilde{y} \right)^{p_j} \mathrm{d}\tilde{\tilde{x}} \\
&\le C((\tilde{B}_j),(p_j)) C((\tilde{\tilde{B}}_j),(p_j)) \\
&\qquad \prod_{j=1}^m \left(\int_{B_jU^\perp}\int_{B_jU} f_j(\tilde{y} + \tilde{\tilde{y}})\,\mathrm{d}\tilde{y}\,\mathrm{d}\tilde{\tilde{y}} \right)^{p_j} \\
&= C((\tilde{B}_j),(p_j)) C((\tilde{\tilde{B}}_j),(p_j)) \prod_{j=1}^m \left(\int_{H_j} f_j(y)\,\mathrm{d}y \right)^{p_j}.
\end{aligned}
$$

Here we have used for the first inequality that for almost any $\tilde{\tilde{x}} \in U^\perp$ the tuple $(f_j(\cdot + B_j\tilde{\tilde{x}}))$ consists of non-negative integrable functions defined on B_jU and we can therefore use the Brascamp–Lieb inequality for the datum $((\tilde{B}_j),(p_j))$. For the next equality we use the definitions of Γ_j and $\tilde{\tilde{B}}_j$ and for the one below that we use the translation invariance of the inner integral and the fact that $\Gamma_j\tilde{\tilde{x}} \in B_jU$ for any $\tilde{\tilde{x}} \in U^\perp$. For the second inequality we use the fact that for any j the inner integral defines a non-negative function of $\tilde{\tilde{B}}_j\tilde{\tilde{x}}$ with domain $(B_jU)^\perp$ and we can therefore use the Brascamp–Lieb inequality for the datum $((\tilde{\tilde{B}}_j),(p_j))$.

Since we can perform this calculation for any tuple of non-negative integrable functions (f_j) defined on H_j, we have established the inequality

$$
C((B_j),(p_j)) \le C((\tilde{B}_j),(p_j)) C((\tilde{\tilde{B}}_j),(p_j)). \tag{2.18}
$$

In particular this shows that if both $C((\tilde{B}_j),(p_j))$ and $C((\tilde{\tilde{B}}_j),(p_j))$ are finite then $C((B_j),(p_j))$ is finite. Since $\dim U < \dim H$ and $\dim U^\perp < H$ we may use the induction hypothesis to establish that this is the case. The positivity condition (2.3) clearly holds since the tuple (p_j) is inherited unchanged from the original datum. The scaling condition (2.2) for $\tilde{B}$ holds by the assumption that U is critical and by subtracting that condition from the scaling condition for H we see that (2.2) holds for $\tilde{\tilde{B}}_j$.

So the only conditions that remain to be checked are (2.1) for any space in $\mathcal{L}_{(\tilde{B}_j)}$ and $\mathcal{L}_{(\tilde{\tilde{B}}_j)}$. First of all, we note that the first of these sets is a subset of $\mathcal{L}_{(B_j)}$.

To see this we note that it is enough to show that the building blocks of $\mathcal{L}_{(\tilde{B}_j)}$, the sets $\ker \tilde{B}_j$, lie in $\mathcal{L}_{(B_j)}$. Since $\tilde{B}_j = B_j|_U$ we get that $\ker \tilde{B}_j = \ker B_j \cap U$ and the inclusion follows as both the sets on the right hand side are elements of $\mathcal{L}_{(B_j)}$. Now, for any $W \in \mathcal{L}_{(\tilde{B}_j)}$ we have that $W \subset U$ and therefore $\dim \tilde{B}_j W = \dim B_j W$. Therefore, the inequality

$$\dim W \leq \sum_j p_j \dim \tilde{B}_j W$$

is in the list in inequalities coming from $\mathcal{L}_{(B_j)}$.

Secondly, we study $\mathcal{L}_{(\tilde{\tilde{B}}_j)}$. Let us take an element $\tilde{\tilde{W}}$ from this set. Our aim is to establish that the inequality

$$\dim \tilde{\tilde{W}} \leq \sum_j p_j \dim \tilde{\tilde{B}}_j \tilde{\tilde{W}}$$

is in the list from $\mathcal{L}_{(B_j)}$. Since U is critical and the elements in the pairs U, W and B_jU, $\tilde{\tilde{B}}_j \tilde{\tilde{W}}$ are orthogonal to each other we see that we may equivalently establish the inequality

$$\dim(\tilde{\tilde{W}} + U) \leq \sum_j p_j \dim(\tilde{\tilde{B}}_j \tilde{\tilde{W}} + B_j U). \tag{2.19}$$

We note that the sets $\tilde{\tilde{B}}_j \tilde{\tilde{W}} + B_j U$ and $B_j(\tilde{\tilde{W}} + U)$ are the same. To see this take an element x from the former set. Then x has the form $\Pi_{(B_jU)^\perp} B_j y + B_j z$ with $y \in \tilde{\tilde{W}}$ and $z \in U$. Now there is an element $y' \in U$ such that $\Pi_{(B_jU)^\perp} B_j y = B_j y + B_j y'$. Then $x = B_j(y + (y' + z))$ with $y \in \tilde{\tilde{W}}$ and $y' + z \in U$. For the other direction we take $x \in B_j(\tilde{\tilde{W}} + U)$. Then we can write $x = B_j(y + z)$ with $y \in \tilde{\tilde{W}}$ and $z \in U$. We take y' as before and then $x = \tilde{\tilde{B}}_j y + B_j(z - y')$ with $y \in \tilde{\tilde{W}}$ and $z - y' \in U$.

Therefore, it is enough to show that $\tilde{\tilde{W}} + U \in \mathcal{L}_{(B_j)}$. To establish this we note first of all that if $\tilde{\tilde{W}} = \ker \tilde{\tilde{B}}_j$ then $\tilde{\tilde{W}} + U = \ker B_j + U$. To see this take $x \in \tilde{\tilde{W}}$. This means by definition that $B_j x \in B_j U$ so $x \in \ker B_j + U$. On the other hand, if we take $x \in \ker B_j$ and write $x = y + z$ with $y \in U$ and $z \in U^\perp$ then $B_j z = B_j x - B_j y = -B_j y \in B_j U$ so $\tilde{\tilde{B}}_j z = 0$ so $z \in \ker \tilde{\tilde{B}}_j$. We also note that for any $\tilde{\tilde{W}}_1, \tilde{\tilde{W}}_2 \in \mathcal{L}_{(\tilde{\tilde{B}}_j)}$ we have that $(\tilde{\tilde{W}}_1 + U) \cap (\tilde{\tilde{W}}_2 + U) = (\tilde{\tilde{W}}_1 \cap \tilde{\tilde{W}}_2) + U$ and $(\tilde{\tilde{W}}_1 + U) + (\tilde{\tilde{W}}_2 + U) = (\tilde{\tilde{W}}_1 + \tilde{\tilde{W}}_2) + U$. The first of those follows from the fact that both $\tilde{\tilde{W}}_1$ and $\tilde{\tilde{W}}_2$ lie in $U^\perp$ and the second is self-evident. Is is now clear that by using induction on the number of operations needed to get to $\tilde{\tilde{W}}$ that we can show that $\tilde{\tilde{W}} + U \in \mathcal{L}_{(B_j)}$ and we thus complete the proof of the theorem. □

By examining the above proof we can give a procedure which tells us when we have found all the conditions included in (2.1).

We start by looking for necessary conditions by going through an enumeration of the elements of $\mathcal{L}_{(B_j)}$ and we decide (arbitrarily) to pause when we have found the necessary conditions (2.1) for $V \in \mathcal{V}$ where $\mathcal{V} \subset \mathcal{L}_{(B_j)}$. At this stage we wish to determine whether we have found all the necessary conditions for the Brascamp–Lieb inequality to hold. The conditions (2.1) for $V \in \mathcal{V}$, together with the conditions (2.2) and (2.3) restrict the set of tuples (p_j) for which the Brascamp–Lieb inequality holds to a polyhedron $\tilde{\mathcal{S}}_{(B_j)}$ and we wish to determine whether $\tilde{\mathcal{S}}_{(B_j)} = \mathcal{S}_{(B_j)}$ where $\mathcal{S}_{(B_j)}$ is the Brascamp–Lieb polyhedron for (B_j). This will be the case if and only if each vertex of $\tilde{\mathcal{S}}_{(B_j)}$ is in $\mathcal{S}_{(B_j)}$. There exists an algorithm which lists all of the vertices of $\tilde{\mathcal{S}}_{(B_j)}$. For each vertex (q_j) in this list we know that m of the conditions (2.1) for $V \in \mathcal{V}$, (2.2) and (2.3) are satisfied with equality. If none of these equalities comes from (2.1) then the support of (q_j) can only contain one element q_{j_0} and we know from above that the Brascamp–Lieb inequality holds at this vertex if and only if $q_{j_0} = 1$ and $\ker B_{j_0} = \{0\}$. Otherwise there is a space $U \in \mathcal{V}$ which lies strictly between $\{0\}$ and H such that (2.1) holds with equality for U. By the proof above we see that the Brascamp–Lieb inequality holds at (q_j) if and only if it holds for the data $((\tilde{B}_j), (q_j))$ and $((\tilde{\tilde{B}}_j), (q_j))$, that is if $(q_j) \in \mathcal{S}_{(\tilde{B}_j)}$ and $(q_j) \in \mathcal{S}_{(\tilde{\tilde{B}}_j)}$.

To determine whether this is the case we run through the above algorithm for both $\mathcal{S}_{(\tilde{B}_j)}$ and $\mathcal{S}_{(\tilde{\tilde{S}}_j)}$. This recursion can only have n levels of depth and will therefore be completed in a finite number of steps and when it is completed we know whether (q_j) is in $\mathcal{S}_{(B_j)}$ in which case we move on to the next vertex, or whether (q_j) is not in $\mathcal{S}_{(B_j)}$ in which case we break the pause and continue looking for necessary conditions in the list of $\mathcal{L}_{(B_j)}$ until we decide again (arbitrarily) to pause and check whether we have now found all of the necessary conditions.

Chapter 3

The Brascamp–Lieb inequality on finite fields

In this short chapter we state and prove an analogue of the Brascamp–Lieb inequality for finite fields. The proof is based on the induction argument in [6] and has therefore many similarities to the argument in Section 2.3.

To set up the problem we let $\mathbb{F}$ be a finite field, m, n and n_j be integers and p_j be non-negative numbers for $j = 1, \dots, m$. Also, let B_j be linear transformations between the vector spaces $\mathbb{F}^n$ and $\mathbb{F}^{n_j}$. The question we pose is: What are necessary and sufficient conditions on this input so that

$$\int_{\mathbb{F}^n} \prod_{j=1}^{m} f_j^{p_j}(B_j x)\,\mathrm{d}x \leq C \prod_{j=1}^{m} \left(\int_{\mathbb{F}^{n_j}} f_j(x)\,\mathrm{d}x \right)^{p_j} \tag{3.1}$$

holds for all non-negative valued functions f_j on $\mathbb{F}^{n_j}$, with a constant C independent of f_j and $|\mathbb{F}|$, the cardinality of $\mathbb{F}$? We take as measure the counting measure on $\mathbb{F}^n$, so that

$$\int_{\mathbb{F}^n} f(x)\,\mathrm{d}x = \sum_{x \in \mathbb{F}^n} f(x).$$

Example 3.1. In the Euclidean case, a basic example which yields restrictions on the input data is to take a subspace V of $\mathbb{R}^n$ and letting f_j be a centred gaussian adapted to V. By this we mean that the set $f_j^{-1}((\frac{1}{2}, \infty))$ (say) is the coin-shaped neighbourhood in $\mathbb{R}^n$ of the unit disk in V. The analogue of this in the finite field case is to take V a subspace of $\mathbb{F}^n$ and simply define

$$f_j(x) = \chi_{B_j V}(x) = \begin{cases} 1 & \text{if } x \in B_j V, \\ 0 & \text{if } x \notin B_j V. \end{cases}$$

We see then that $\prod_{j=1}^{m} f_j^{p_j}(B_j x) = 1$ for all $x \in V$, so plugging into (3.1) we get

$$|\mathbb{F}|^{\dim V} \leq C |\mathbb{F}|^{\sum_{j=1}^{m} p_j \dim(B_j V)}$$

and we need this to hold independently of $\mathbb{F}$ so we get the condition

$$\dim V \leq \sum_{j=1}^{m} p_j \dim(B_j V) \tag{3.2}$$

for all subspaces V of $\mathbb{F}^n$. Furthermore, the particular case $V = \{0\}$ gives that $C \geq 1$.

This is in fact the only condition that needs to be satisfied for (3.1) to hold with a constant independent of $|\mathbb{F}|$. The theorem can be stated as follows:

Theorem 3.2. *Let $\mathbb{F}$, n, m, n_j, p_j and B_j be as before and assume that*

$$p_j \geq 0 \qquad \textit{for all } j = 1, \ldots, m \textit{ and} \tag{3.3}$$

$$\dim V \leq \sum_{j=1}^{m} p_j \dim(B_j V) \qquad \textit{for all subspaces } V \textit{ of } \mathbb{F}^n. \tag{3.4}$$

Then

$$\int_{\mathbb{F}^n} \prod_{j=1}^{m} f_j^{p_j}(B_j x)\,\mathrm{d}x \leq C \prod_{j=1}^{m} \left(\int_{\mathbb{F}^{n_j}} f_j(x)\,\mathrm{d}x \right)^{p_j} \tag{3.5}$$

holds with the constant $C = 1$ and this constant is the best possible. Conversely, if this equation holds with any constant C, independent of $\mathbb{F}$, then conditions (3.3) *and* (3.4) *must be satisfied.*

Proof. We have already noted the example that establishes the latter claim.

For the other part we use induction on $n + m$. For the base case we assume $m = 1$. By testing (3.4) on $\ker B_1$ we get that B_1 must be surjective. Then testing (3.4) on $\mathbb{F}^n$ we get that $p_1 \geq 1$ so the result we want to prove is contained in the inequality

$$\int_{\mathbb{F}^n} f^p(x)\,\mathrm{d}x \leq \left(\int_{\mathbb{F}^n} f(x)\,\mathrm{d}x \right)^p \tag{3.6}$$

which holds for all non-negative functions f and is just expressing the embedding of ℓ^p in ℓ^1 for $p \geq 1$.

For the inductive step in the proof of the theorem we consider four cases:

1. $p_{j_0} = 0$ for some j_0.

   If f_{j_0} is identically 0 then there is nothing to prove. Otherwise

   $$\int_{\mathbb{F}^n} \prod_{j=1}^{m} f_j^{p_j}(B_j x)\,\mathrm{d}x = \int_{\mathbb{F}^n} \prod_{\substack{j=1 \\ j \neq j_0}}^{m} f_j^{p_j}(B_j x)\,\mathrm{d}x$$

   and

   $$\prod_{j=1}^{m} \left(\int_{\mathbb{F}^{n_j}} f_j(x)\,\mathrm{d}x \right)^{p_j} = \prod_{\substack{j=1 \\ j \neq j_0}}^{m} \left(\int_{\mathbb{F}^{n_j}} f_j(x)\,\mathrm{d}x \right)^{p_j}$$

   so the result follows from the induction hypothesis.

2. $\dim U = \sum_{j=1}^{m} p_j \dim(B_jU)$ for some proper subspace U of $\mathbb{F}^n$ where by a proper subspace we mean that $U \notin \{\{0\}, \mathbb{F}^n\}$.

Let us say that $r = \dim U$ and take a basis $\{u_1, \ldots, u_r\}$ for U. By Steiner's theorem we can extend this to a basis $\{u_1, \ldots, u_n\}$ for $\mathbb{F}^n$. We denote by $\tilde{U}$ the subspace of $\mathbb{F}^n$ spanned by $u_{r+1}, \ldots, u_n$. Then we can write the left hand side of (3.5) as

$$\int_{\mathbb{F}^n} \prod_{j=1}^{m} f_j^{p_j}(B_jx)\,\mathrm{d}x = \int_{\tilde{U}} \int_{U} \prod_{j=1}^{m} f_j^{p_j}(\tilde{B}_j\tilde{x} + B_j\tilde{\tilde{x}})\,\mathrm{d}\tilde{x}\,\mathrm{d}\tilde{\tilde{x}}$$

where $x = \tilde{x} + \tilde{\tilde{x}}$ with $\tilde{x} \in V$ and $\tilde{\tilde{x}} \in \tilde{U}$ and $\tilde{B}_j : U \to B_jU$ is the restriction of B_j. From condition (3.4) we then get that

$$\dim V \leq \sum_{j=1}^{m} p_j \dim(\tilde{B}_jV)$$

for any subspace V of U and we can apply the Brascamp–Lieb inequality to the tuple $(f_j(\cdot + B_j\tilde{\tilde{x}}))$ with the datum $((\tilde{B}_j), (p_j))$ which we know to hold by the induction hypothesis. This allows us to estimate the right hand side of the previous expression by

$$\int_{\tilde{U}} \prod_{j=1}^{m} \left(\int_{B_jU} f_j(\tilde{y} + B_j\tilde{\tilde{x}})\,\mathrm{d}\tilde{y} \right)^{p_j} \mathrm{d}\tilde{\tilde{x}} \tag{3.7}$$

For each $\tilde{\tilde{x}} \in \tilde{U}$ we can write $B_j\tilde{\tilde{x}} = \tilde{\tilde{B}}_j\tilde{\tilde{x}} + \Gamma_j\tilde{\tilde{x}}$ where $\tilde{\tilde{B}}_j\tilde{\tilde{x}} \in \widetilde{B_jU}$ and $\Gamma_j\tilde{\tilde{x}} \in B_jU$ where $\widetilde{B_jU}$ is a complementary subspace to B_jU in H_j. Thus by making a translation in the $\tilde{x}$ variable we see that we can write the inner integrals in the previous expression as

$$\left(\int_{B_jU} f_j(\tilde{y} + \tilde{\tilde{B}}_j\tilde{\tilde{x}})\,\mathrm{d}\tilde{y} \right).$$

Similarly to the previous chapter we can show that condition (3.4) holds for the datum $((\tilde{\tilde{B}}_j), (p_j))$ since it holds for $((B_j), (p_j))$ and there is equality in the condition for U. Therefore, by the induction hypothesis, we can estimate the quantity in (3.7) by

$$\prod_{j=1}^{m} \left(\int_{\widetilde{B_jU}} \int_{B_jU} f_j(\tilde{y} + \tilde{\tilde{y}})\,\mathrm{d}\tilde{y}\,\mathrm{d}\tilde{\tilde{y}} \right)^{p_j} = \prod_{j=1}^{m} \left(\int_{B_j\mathbb{F}^n} f_j(y)\,\mathrm{d}y \right)^{p_j}$$

and this is the estimate we required.

3. $n = \sum_{j=1}^{m} p_j \dim(B_j\mathbb{F}^n)$, but $\dim V < \sum_{j=1}^{m} p_j \dim(B_jV)$ for all proper subspaces V of $\mathbb{F}^n$ and $p_j > 0$ for all j.

   We have already dealt with the possibility $m = 1$ in the base case so we may assume $m \geq 2$. Then we note that the conditions we are studying require the vector $\bar{p} = (p_1, \dots, p_m)$ to lie in an open subset P of the hyperplane in $\mathbb{R}^m$ defined by $n = \sum_{j=1}^{m} p_j \dim(B_j\mathbb{F}^n)$. In particular, if we let

$$\tilde{p}_{1,t} = p_1 + t\dim(B_2\mathbb{F}^n) \quad \text{and}$$
$$\tilde{p}_{2,t} = p_1 - t\dim(B_1\mathbb{F}^n)$$

   then $\tilde{p}_t = (\tilde{p}_{1,t}, \tilde{p}_{2,t}, p_3, \dots, p_m)$ lie in this same hyperplane for all $t \in \mathbb{R}$ and by the openness of P we see that there exists a $t > 0$ such that $\tilde{p}_t \in P$. However, the set of such t's is bounded above by $\frac{p_2}{\dim(B_1\mathbb{F}^n)}$ so by the least upper bound axiom and the openness of P there is a $t^+ > 0$ such that $\tilde{p}_t \in P$ for all $0 < t < t^+$ but $\tilde{p}_{t^+} \notin P$. In fact we must either have $\tilde{p}_{2,t^+} = 0$ or $\dim V = \sum_{j=1}^{m} p_j \dim(B_jV)$ for some subspace V of $\mathbb{F}^n$. Then we can use the arguments from either case 1 or 2 to establish that (3.5) holds for $\bar{p}_{t^+}$.

   The same argument gives that there exists a $t^- < 0$ such that (3.5) holds for $\bar{p}_{t^-}$ and then we can use interpolation to see that (3.5) holds at $\bar{p}$.

4. We have $p_j > 0$ for all j and $\dim V < \sum_{j=1}^{m} p_j \dim(B_jV)$ for all subspaces V of $\mathbb{F}^n$, including $\mathbb{F}^n$ itself.

   In a similar manner as in step 3 we can find a real number q_1 such that $p_1 > q_1 \geq 0$ and such that the hypothesis for this case continues to hold with p_1 replaced by any number strictly between p_1 and q_1 but fails with p_1 replaced by any number smaller than q_1. Again by an openness argument we see that at least one of the inequalities (3.3) or (3.4) becomes an equality with p_1 replaced by q_1. Now, one of the previous three cases gives (3.5) with q_1 instead of p_1 and then we can do the following calculation:

$$\begin{aligned}
\int_{\mathbb{F}^n} \prod_{j=1}^{m} f_j^{p_j}(B_jx)\,\mathrm{d}x &= \int_{\mathbb{F}^n} (f_1^{\frac{p_1}{q_1}})^{q_1}(B_1x) \prod_{j=2}^{m} f_j^{p_j}(B_jx)\,\mathrm{d}x \\
&\leq \left(\int_{\mathbb{F}^{n_1}} f_1^{\frac{p_1}{q_1}}(B_1x)\,\mathrm{d}x\right)^{q_1} \prod_{j=2}^{m} \left(\int_{\mathbb{F}^{n_j}} f_j(B_jx)\,\mathrm{d}x\right)^{p_j} \\
&\leq \prod_{j=1}^{m} \left(\int_{\mathbb{F}^{n_j}} f_j(B_jx)\,\mathrm{d}x\right)^{p_j}
\end{aligned}$$

   where we have used (3.6) in the last step.

This completes the proof of Theorem 3.2 by induction. □

Example 3.1 suggests functions which play the role that gaussians play in the Euclidean case, namely characteristic functions of subspaces of $\mathbb{F}$. We can make this analogy stronger and prove a version of Lieb's theorem for finite fields. (This proof was done in close collaboration with Tony Carbery.)

We approach this theorem via a convolution inequality going back to Ball. Fix (B_j) and (p_j) and let $\bar{C}$ denote best constant for (3.1) and $C(f_j)$ denote the constant which the tuple (f_j) attains. Then

$$\begin{aligned}
C(f_j)C(f_j')\prod_{j=1}^m\left(\int f_j\right)^{p_j}\left(\int f_j'\right)^{p_j} &= \int\prod_{j=1}^m f_j^{p_j}(B_jy)\,\mathrm{d}y\int\prod_{j=1}^m f_j'^{p_j}(B_jx)\,\mathrm{d}x\\
&= \int\int\prod_{j=1}^m f_j^{p_j}(B_jy)f_j'^{p_j}(B_jx-B_jy)\,\mathrm{d}y\,\mathrm{d}x\\
&\le \bar{C}\int\prod_{j=1}^m\left(\int f_j(y)f_j'(B_jx-y)\,\mathrm{d}y\right)^{p_j}\mathrm{d}x\\
&= \bar{C}\int\prod_{j=1}^m(f_j*f_j')^{p_j}(B_jx)\,\mathrm{d}x\\
&= \bar{C}C(f_j*f_j')\prod_{j=1}^m\left(\int f_j*f_j'\right)^{p_j}
\end{aligned}$$

where we have used (3.1) on $(f_j(\cdot)f_j'(B_jx-\cdot))$ for the inequality. By cancelling the integrals which is in order provided that none of f_j's and f_j''s is identically zero we get that

$$\frac{C(f_j)}{\bar{C}}\frac{C(f_j')}{\bar{C}}\le\frac{C(f_j*f_j')}{\bar{C}}. \tag{3.8}$$

In particular we note that if (f_j) and (f_j') are optimisers for (3.1) then so is (f_j*f_j').

Now, inequality (3.1) is exhausted by tuples (f_j) such that $\int f_j=1$ for all j. These tuples form a compact set and the existence of an optimiser (f_j) of this form is therefore guaranteed. Now let $f_j^{(\nu)}$ denote the ν-fold convolution of f_j with itself. Then by (3.8) we have that $(f_j^{(\nu)})$ is also an optimiser and we note that $\int f_j^{(\nu)}=1$ for all j and ν. Let V_j be the smallest subspace of $\mathbb{F}^{n_j}$ for which there exists an $x_{0j}\in\mathbb{F}^{n_j}$ such that V_j+x_{0j} contains the support of f_j.

We now introduce the Fourier transform of f_j as follows. Let $\omega:\mathbb{F}\to S^1$ be a character on $\mathbb{F}$. Then for $\xi\in\mathbb{F}^{n_j*}$ we let

$$\hat{f}_j(\xi)=\int_{\mathbb{F}^{n_j}}f_j(x)\omega(\xi x)\,\mathrm{d}x$$

where as usual ξx denotes the action of the linear map ξ on x. Now let $\{v_1,\dots,v_{k_j}\}$ be a basis for V_j and extend it to a basis $\{v_1,\dots,v_{n_j}\}$ for $\mathbb{F}^{n_j}$ and let $\{v_1^*,\dots,v_{n_j}^*\}$

be the dual basis for $\mathbb{F}^{n_j*}$. Then for any ξ in $\tilde{V}_j^*$, the span of $\{v_{k_j+1}^*, \dots, v_{n_j}^*\}$, we have that

$$\begin{aligned}\hat{f}_j(\xi) = & \int_{V_j + x_{0j}} f_j(x)\omega(\xi x)\,\mathrm{d}x = \omega(\xi x_{0j}) \int_{V_j} f_j(x + x_{0j})\omega(\xi x)\,\mathrm{d}x \\ =& \omega(\xi x_{0j}) \int f_j(x)\,\mathrm{d}x = \omega(\xi x_{0j})\end{aligned}$$

so in particular $|\hat{f}_j(\xi)| = 1$.

On the other hand we have that $|\hat{f}_j(\xi)| \le 1$ for all $\xi \in \mathbb{F}^{n_j*}$ and if $|\hat{f}_j(\xi)| = 1$ for some ξ then $\xi x = \xi x'$ for any x and x' such that $x + x_{0j}$ and $x' + x_{0j}$ are in the support of f_j. Let us see that this guarantees that $\xi x = 0$ for any $x \in V_j$. If this is not the case, there exists a proper subspace U of V_j such that $\xi x \neq 0$ for any $x \in V_j \setminus U$. Without loss of generality, we may assume that U is spanned by $\{v_2, \dots, v_{k_j}\}$. Let x_i be the coefficient of v_i in the representation of x. If x_1 is the same for all $x \in \operatorname{supp} f_j$ then the support of f_j is contained in a translation of U, contrary to the assumption that V is the smallest space with that property. Therefore, there exist $x, x' \in V$ such that $x_1 \neq x_1'$ and such that $x + x_{0j}$ and $x' + x_{0j}$ are in the support of f_j. Then, by what we said above, we get that $\xi(x - x') = 0$ whereas $x - x' \notin U$. This contradicts the definition of U so $\xi x = 0$ for any $x \in V_j$ and therefore $\xi \in \tilde{V}_j^*$.

We have that $(f_j * f_j')^\wedge(\xi) = \hat{f}_j(\xi)\hat{f}_j'(\xi)$ so if we let $\bar{f}_j = f_j^{(\nu)}$ with $\nu = |\mathbb{F}|$ then we see that $\hat{\bar{f}}(\xi) = 1$ for all $\xi \in \tilde{V}_j^*$ and $\hat{\bar{f}}(\xi) < 1$ for all $\xi \notin \tilde{V}_j^*$. From this we see that $\hat{\bar{f}}^{(\nu)}$ tends to $\chi_{\tilde{V}_j^*}$, the characteristic function of $\tilde{V}_j^*$, as ν tends to infinity. It is a simple calculation to verify that $\hat{\chi}_{V_j} = \chi_{\tilde{V}_j^*}$ and this shows that $\bar{f}_j^{(\nu)}$ tends to χ_{V_j} as ν tends to infinity.

We have thus proved the following.

Theorem 3.3 (Lieb's theorem for finite fields). *Let a Brascamp–Lieb datum* $((B_j), (p_j))$ *be given. Then there exists an extremiser* (f_j) *for* (3.1) *such that* f_j *is a characteristic function of a subspace of* $\mathbb{F}^{n_j}$.

We remark that this theorem holds independently of the subspace condition (3.4), so it holds whether or not the best constant in (3.1) depends on the cardinality of the underlying field.

Chapter 4

Heat flow and optimisers for Brascamp–Lieb inequalities

4.1 Monotonicity properties

In this chapter we discuss the heat flow methods that have been developed to derive monotonicity formulas for quantities related to the Brascamp–Lieb inequality. We present a formula for the monotonicity of heat flow for the general Brascamp–Lieb problem and use it to give the form of the optimisers for any Brascamp–Lieb datum.

To set the problem up, let f_j be non-negative integrable functions on H for $j = 1, \ldots, m$. Let $f_j(x,t)$ for $t \geq 0$ be the solution to the initial value problem

$$\begin{aligned} f_j(x,0) &= f_j(x) \\ \frac{\partial}{\partial t} f_j(x,t) &= \Delta_{A_j} f_j(x,t) \qquad t > 0 \end{aligned} \tag{4.1}$$

where $\Delta_{A_j} f_j = \operatorname{div}(A_j \nabla f_j)$ and A_j is a positive definite linear transformation on H.

Let us define the product

$$F(x,t) = \prod_{j=1}^{m} f_j^{p_j}(x,t).$$

Our aim is to discuss monotonicity properties of $\int F$, that is, under what circumstances we can say that

$$\frac{\partial}{\partial t} \int F(x,t)\,\mathrm{d}x \geq 0.$$

From the definitions we see that

$$\frac{\partial}{\partial t} F = F \sum_{j=1}^{m} p_j \frac{\Delta_{A_j} f_j}{f_j} = F \sum_{j=1}^{m} p_j \frac{\operatorname{div}(v_j f_j)}{f_j}$$

where $v_j = \frac{A_j \nabla f_j}{f_j}$ so if we let $h_j = \log f_j$ then $v_j = A_j \nabla h_j$. We calculate further

$$\frac{\partial}{\partial t} F = F \sum_{j=1}^{m} p_j \left(\operatorname{div}(v_j) + \langle \frac{\nabla f_j}{f_j}, v_j \rangle \right) = I + II.$$

Note that

$$\int II = \int F \sum_{j=1}^{m} p_j \langle \nabla h_j, A_j \nabla h_j \rangle.$$

In the literature concerning this problem on $\mathbb{R}^n$ and S^n there have been several ideas put forward on how to proceed from this point. In [7] the authors proceed on $\mathbb{R}^n$ by using the log-concavity of the heat kernel. Another approach which appears in [16] is to use integration by parts for I and write

$$\begin{aligned}\int I &= \int F \sum_{j=1}^{m} p_j \operatorname{div}(v_j) = \sum_{j=1}^{m} p_j \int F \operatorname{div}(v_j) \\ &= \sum_{j=1}^{m} p_j \left(\int \operatorname{div}(F v_j) - \int \langle \nabla F, v_j \rangle \right) = - \int F \langle \sum_{j'=1}^{m} p_{j'} \nabla h_{j'}, \sum_{j=1}^{m} p_j A_j \nabla h_j \rangle\end{aligned}$$

where we have used the divergence theorem to eliminate the first term in the next to last expression. This is justified for any $t > 0$ since H is boundaryless and the integrand has enough smoothness.

Let us now assume $f_j(x) = \tilde{f}_j(B_j x)$ where $\tilde{f}_j$ is a positive integrable function on H_j and $B_j : H \to H_j$ is a linear surjective map such that $B_j B_j^* = \mathrm{id}_{H_j}$. This assumption entails that B_j acts as the orthogonal projection from H onto the orthogonal complement of $\ker B_j$. When it is convenient we will use this. Then we have in particular that $H_j \subset H$. (We note that this condition on B_j does not restrict the applicability of the results as any surjective linear map $H \to H_j$ can be written as a composition of such a B_j and an invertible linear map from H_j to itself.)

With these assumptions let us show that

$$f_j(x,t) = \int_H \frac{\det A_j^{\frac{1}{2}}}{(4\pi t)^{\frac{n}{2}}} e^{-\frac{\langle x-y, A_j(x-y)\rangle}{4t}} \tilde{f}_j(B_j y) \, \mathrm{d}y.$$

is a solution to (4.1). It is a straightforward calculation which shows that the relationship between the derivatives holds so the only thing we need to check is that $f_j(x,0) = f_j(x)$. Since we may assume that we have the decomposition $H = H_j \oplus H_j^{\perp}$ we can write $f_j(x,t)$ as

$$\int_{H_j} \int_{H_j^{\perp}} \frac{\det A_j^{\frac{1}{2}}}{(4\pi t)^{\frac{n}{2}}} \exp\left(-\frac{1}{4t} \left\langle \begin{pmatrix} B_j x - y_1 \\ B_j^{\perp} x - y_2 \end{pmatrix}, A_j \begin{pmatrix} B_j x - y_1 \\ B_j^{\perp} x - y_2 \end{pmatrix} \right\rangle \right) \tilde{f}_j(y_1) \, \mathrm{d}y_2 \, \mathrm{d}y_1$$

where $B_j^\perp$ denotes the projection onto the orthogonal complement of H_j. In the inner integral we make the change of variables $y_2 \mapsto y_2 - B_j^\perp x$ and get

$$\int_{H_j}\left(\int_{H_j^\perp}\frac{\det A_j^{\frac{1}{2}}}{(4\pi t)^{\frac{n}{2}}}\exp\left(-\frac{1}{4t}\left\langle\left(\begin{smallmatrix}B_jx-y_1\\ -y_2\end{smallmatrix}\right), A_j\left(\begin{smallmatrix}B_jx-y_1\\ -y_2\end{smallmatrix}\right)\right\rangle\right)\mathrm{d}y_2\right)\tilde{f}_j(y_1)\,\mathrm{d}y_1. \quad (4.2)$$

Let us write A_j in blocks corresponding to the decomposition of H as $H_j \oplus H_j^\perp$, so

$$A_j = \begin{pmatrix} A_{1j} & A_{2j} \\ A_{2j}^* & A_{3j}\end{pmatrix}.$$

Then the inner product in the exponent above can be written as

$$\begin{aligned}\langle A_{1j}(B_jx-y_1), B_jx-y_1\rangle - 2\langle B_jx-y_1, A_{2j}y_2\rangle + \langle y_2, A_{3j}y_2\rangle \\ = \langle (A_{1j} - A_{2j}A_{3j}^{-1}A_{2j}^*)(B_jx-y_1), B_jx-y_1\rangle \\ +\langle A_{3j}^{\frac{1}{2}}y_2 - A_{3j}^{-\frac{1}{2}}A_{2j}^*(B_jx-y_1), A_{3j}^{\frac{1}{2}}y_2 - A_{3j}^{-\frac{1}{2}}A_{2j}^*(B_jx-y_1)\rangle\end{aligned}$$

and if we now carry out the y_2 integration in (4.2) we get

$$\int_{H_j}\left(\frac{\det \tilde{A}_j^{\frac{1}{2}}}{(4\pi t)^{\frac{n_j}{2}}}e^{-\frac{1}{4t}\langle \tilde{A}_j(B_jx-y_1), B_jx-y_1\rangle}\right)\tilde{f}_j(y_1)\,\mathrm{d}y_1. \quad (4.3)$$

where we have written $\tilde{A}_j = A_{1j} - A_{2j}A_{3j}^{-1}A_{2j}^*$ and used that $\det\tilde{A}_j \det A_{3j} = \det A_j$.

Let us call the quantity within the parentheses $K_{j,t}(B_jx - y_1)$. Then the whole integral is $(K_{j,t} * \tilde{f}_j)(B_jx)$. Now we note that $K_{j,t}$ is an approximation to the identity on H_j and $\tilde{f}_j$ is integrable so we see that $\lim_{t\to 0}(K_{j,t} * \tilde{f}_j)(B_jx) = \tilde{f}_j(B_jx) = f_j(x)$. This confirms that $f_j(x,t)$ is a solution to (4.1). Furthermore, we see that $f_j(x,t)$ depends only on B_jx so we can write $f_j(x,t)$ as $\tilde{f}_j(B_jx,t)$.

We will now calculate the limit of $\int F(x,t)\,\mathrm{d}x$ as t tends to infinity. We have that

$$\begin{aligned}\int_H \prod_{j=1}^m f_j^{p_j}(x,t)\,\mathrm{d}x &= \int_H \prod_{j=1}^m f_j^{p_j}(t^{\frac{1}{2}}w,t)t^{\frac{n}{2}}\,\mathrm{d}w \\ &= \int_H \prod_{j=1}^m \left(t^{\frac{n_j}{2}}f_j(t^{\frac{1}{2}}w,t)\right)^{p_j}\mathrm{d}w\end{aligned}$$

where we have made the change of variables $x = t^{\frac{1}{2}}w$ and used the necessary condition (1.20). From (4.3) we see that we can write the quantity within the parentheses as

$$\int_{H_j}\frac{\det\tilde{A}_j^{\frac{1}{2}}}{(4\pi)^{\frac{n_j}{2}}}e^{-\frac{1}{4}\left\langle \tilde{A}_j(B_jw - t^{-\frac{1}{2}}y_1), B_jw - t^{-\frac{1}{2}}y_1\right\rangle}\tilde{f}_j(y_1)\,\mathrm{d}y_1.$$

Thus by dominated convergence which is applicable as $\cap_j \ker B_j = \{0\}$ since the datum is gaussian extremisable we get that

$$\lim_{t\to\infty} \int_H \prod_{j=1}^m f_j^{p_j}(x,t)\,\mathrm{d}x = \int_H \prod_{j=1}^m L_j^{p_j}(x)\,\mathrm{d}x \prod_{j=1}^m \left(\int_{H_j} \tilde{f}_j\right)^{p_j}$$

where

$$L_j(x) = \frac{\det \tilde{A}_j^{\frac{1}{2}}}{(4\pi)^{\frac{n_j}{2}}} e^{-\frac{1}{4}\langle \tilde{A}_j B_j w, B_j w\rangle}.$$

We thus conclude that

$$\lim_{t\to\infty} \int_H \prod_{j=1}^m f_j^{p_j}(x,t)\,\mathrm{d}x = \left(\frac{\prod_j \det \tilde{A}_j^{p_j}}{\det \sum_j p_j B_j^* \tilde{A}_j B_j}\right)^{\frac{1}{2}} \prod_{j=1}^m \left(\int_{H_j} \tilde{f}_j\right)^{p_j}. \tag{4.4}$$

In what follows we shall fix a time $t > 0$ and mostly not write it explicitly in the equations. We get that

$$\nabla h_j(x) = \frac{\nabla(\tilde{f}_j \circ B_j)(x)}{(\tilde{f}_j \circ B_j)(x)} = \left(\frac{D\tilde{f}_j(B_j x)}{\tilde{f}_j(B_j x)} B_j x\right)^* = (Dk_j(B_j x) B_j x)^*$$

where $k_j = \log \tilde{f}_j$. We will sometimes place the point at which a derivative is evaluated in a subscript; with this notation the last expression becomes $((D_{B_j x} k_j) B_j x)^*$.

We have that

$$\frac{\partial}{\partial t} \int F(x,t)\,\mathrm{d}x = \int F(x,t)\Xi(x,t)\,\mathrm{d}x$$

where

$$\begin{aligned}\Xi(x) = &\sum_{j=1}^m p_j \langle ((D_{B_j x} k_j) B_j)^*, A_j ((D_{B_j x} k_j) B_j)^* \rangle \\ &- \langle \sum_{j'=1}^m p_{j'} ((D_{B_j x} k_j) B_j)^*, \sum_{j=1}^m p_j A_j ((D_{B_j x} k_j) B_j)^* \rangle\end{aligned}$$

and from this we see that we would like to choose A_j so that this quantity is non-negative because this would make the integrand $F\Xi$ pointwise non-negative. We will be able to take all of the A_j's to be the same and we will write $\langle v, A_j w\rangle = v^* R w$.

The quantity above becomes

$$\sum_{j=1}^m p_j (D_{B_j x} k_j) B_j R B_j^* (D_{B_j x} k_j)^* - \sum_{j=1}^m \sum_{j'=1}^m p_j p_{j'} (D_{B_j x} k_{j'}) B_{j'} R B_j^* (D_{B_j x} k_j)^*. \tag{4.5}$$

To get a handle on this we follow [16] but for the most part we choose to express the method in terms of the quantities that are introduced in [7].

So, let us assume that the Brascamp–Lieb datum $((B_j),(p_j))$ is gaussian extremisable. We take a gaussian extremiser (S_j) and recall from [7] that for each j, S_j is an $n_j \times n_j$ symmetric positive definite linear transformation such that (S_j) is a solution to

$$M = \sum_{j=1}^{m} p_j B_j^* S_j B_j \tag{4.6}$$

$$S_j^{-1} = B_j M^{-1} B_j^* \quad j = 1, \ldots, m. \tag{4.7}$$

If we define $\tilde{S}_j := p_j S_j$ we can write this set of equations as

$$M = \sum_{j=1}^{m} (\tilde{S}_j^{\frac{1}{2}} B_j)^* (\tilde{S}_j^{\frac{1}{2}} B_j) \tag{4.8}$$

$$p_j I = (\tilde{S}_j^{\frac{1}{2}} B_j) M^{-1} (\tilde{S}_j^{\frac{1}{2}} B_j)^* \quad j = 1, \ldots, m \tag{4.9}$$

where I is the identity transformation (on H_j). We note that $M = T^*T$ where

$$T = \begin{pmatrix} --- & \tilde{S}_1^{\frac{1}{2}} B_1 & --- \\ & \vdots & \\ --- & \tilde{S}_m^{\frac{1}{2}} B_m & --- \end{pmatrix}.$$

Define $\tilde{m} := \sum_{j=1}^{m} n_j$. Let E_j be such that $E_j T = \tilde{S}_j^{\frac{1}{2}} B_j$, that is E_j is the $n_j \times \tilde{m}$ matrix so that the operator $T \mapsto E_j T$ reads off the relevant n_j lines of T. If we let $P := TM^{-1}T^*$ then we can interpret the second equation for $j = 1$ as saying that the $n_1 \times n_1$ submatrix in the top left corner of P is p_1 times the identity matrix. We can then make a similar statement for all the j's. Finally, we remark that the definition of P in terms of T shows that P is a projection transformation.

If we now take $R = M^{-1}$ and use these definitions we can write (4.5) as

$$\sum_{j=1}^{m} p_j (D_{B_j x} k_j) \tilde{S}_j^{-\frac{1}{2}} \tilde{S}_j^{\frac{1}{2}} B_j M^{-1} B_j^* \tilde{S}_j^{*\frac{1}{2}} \tilde{S}_j^{*-\frac{1}{2}} (D_{B_j x} k_j)^*$$
$$- \sum_{j=1}^{m} \sum_{j'=1}^{m} p_j p_{j'} (D_{B_j x} k_{j'}) \tilde{S}_{j'}^{-\frac{1}{2}} \tilde{S}_{j'}^{\frac{1}{2}} B_{j'} M^{-1} B_j^* \tilde{S}_j^{*\frac{1}{2}} \tilde{S}_j^{*-\frac{1}{2}} (D_{B_j x} k_j)^*.$$

We use (4.9) on the first term and the definition of P on the second to write this expression as

$$\sum_{j=1}^{m} p_j^2 (D_{B_j x} k_j) \tilde{S}_j^{-\frac{1}{2}} E_j I E_j^* \tilde{S}_j^{*-\frac{1}{2}} (D_{B_j x} k_j)^*$$
$$- \sum_{j=1}^{m} \sum_{j'=1}^{m} p_j p_{j'} (D_{B_{j'} x} k_{j'}) \tilde{S}_{j'}^{-\frac{1}{2}} E_{j'} P E_j^* \tilde{S}_j^{*-\frac{1}{2}} (D_{B_j x} k_j)^*$$

where I is the identity matrix of dimension $\tilde{m}$. Since $E_j I E_{j'}^* = 0$ if $j \neq j'$ we can write this more compactly as

$$\Xi = K(x)(I - P)K^*(x). \tag{4.10}$$

Here $K(x)$ is the $1 \times \tilde{m}$ matrix given by

$$K(x) := \left(p_1 (D_{B_1 x} k_1) \tilde{S}_1^{-\frac{1}{2}}, \ldots, p_m (D_{B_m x} k_m) \tilde{S}_m^{-\frac{1}{2}} \right).$$

Since P is a projection transformation it is clear that $\Xi \geq 0$ and we have therefore shown that

$$\frac{\partial}{\partial t} \int F(x,t)\, \mathrm{d}x \geq 0 \tag{4.11}$$

for all $t > 0$.

We have thus established that $\int F(x, t_1) \leq \int F(x, t_2)$ for any $0 < t_1 \leq t_2$. We know from the discussion above that

$$\lim_{t \to 0} F(x,t) = \prod_{j=1}^{m} \tilde{f}_j^{p_j}(B_j x)$$

so by Fatou's lemma we get that

$$\int_H \prod_{j=1}^{m} \tilde{f}_j^{p_j}(B_j x)\, \mathrm{d}x = \int_H F(x,0)\, \mathrm{d}x \leq \int_H F(x,t)\, \mathrm{d}x$$

for all $t > 0$.

By comparing the limits as t tends to zero and infinity we thus arrive at the inequality

$$\int_H \prod_{j=1}^{m} \tilde{f}_j^{p_j}(B_j x)\, \mathrm{d}x \leq \left(\frac{\prod_j \det \tilde{A}_j^{p_j}}{\det \sum_j p_j B_j^* \tilde{A}_j B_j} \right)^{\frac{1}{2}} \prod_{j=1}^{m} \left(\int_{H_j} \tilde{f}_j \right)^{p_j}. \tag{4.12}$$

4.2 Determination of optimisers

Let us now use this set-up to determine all optimisers for the Brascamp–Lieb inequality. This is a problem that has been settled in the rank one case in [3] and [16]. In [7] an analysis of the general case is started but they "cannot fully generalise [...] to the higher rank case".

In the first instance we shall assume that we are working with a geometric datum but it is straightforward to pass from that case to the general one and we shall do so in Theorem 4.8 at the end of this section.

By the condition imposed on B_j we may assume that H_j, the image of B_j, is a subspace of H and that B_j is the orthogonal projection from H onto H_j.

Also, we will take $B_j^{\perp}$ to be the orthogonal projection from H onto $H_j^{\perp}$ the orthogonal complement of H_j. We recall from [7] and the Introduction that a pair of subspaces (V, W) of H is said to be a critical pair if neither V nor W is H, if they are complementary in the sense that $V \cap W = \{0\}$ and $V + W = H$ and if the pair (B_jU, B_jW) is complementary in H_j for each j. Then it is a theorem from [7] that both U and W are critical in the sense that there is equality in (2.1) for them. Furthermore, if the datum is geometric then we may take V and W to be orthogonal complements. We can repeat this splitting until we arrive at a *maximal critical decomposition* where we write H as a sum of pairwise orthogonal spaces, each of which is critical and has no critical subspace.

We now make the following definition:

Definition 4.1. A subspace $\tilde{H}$ of H will be said to be *independent* with respect to the geometric datum $((B_j), (p_j))$ if it is not $\{0\}$ and has the form

$$\tilde{H} = \bigcap_{j=1}^{m} H_j^a$$

where for each j, H_j^a is either H_j or $H_j^{\perp}$.

Clearly there are at most 2^m independent subspaces for any datum and any two distinct independent subspaces are orthogonal to one another. The following is then a sensible definition.

Definition 4.2. The *independent decomposition* of H is the decomposition

$$H = \tilde{\tilde{H}}_1 \bigoplus \tilde{\tilde{H}}_2 = \left(\bigoplus_{k=1}^{k_0} \tilde{H}_k \right) \bigoplus \tilde{\tilde{H}}_2$$

where $\{\tilde{H}_k | k = 1, \ldots, k_0\}$ is an enumeration of the independent subspaces of H and $\tilde{\tilde{H}}_2$ is the orthogonal complement of $\tilde{\tilde{H}}_1$.

The following lemma establishes the relationship between the concepts of criticality and independence.

Lemma 4.3. *Let $\tilde{H}$ be an independent subspace of H and V be a critical one. Then*

1. *$\tilde{H}$ is also critical and*

2. *V can be decomposed as the direct sum $V = (V \cap \tilde{H}) \oplus (V \cap \tilde{H}^{\perp})$ and these two spaces are critical if they are not $\{0\}$.*

Proof. We prove the first part by showing that $(\tilde{H}, \tilde{H}^\perp)$ is a critical pair. For any B_j there are two possibilities, either $\tilde{H} \subset H_j$ or $\tilde{H} \subset H_j^\perp$. If $\tilde{H} \subset H_j$ then we can write $\tilde{H}^\perp$ as the orthogonal sum of $\tilde{\tilde{H}}$ and $H_j^\perp$ where $\tilde{\tilde{H}}$ is the orthogonal complement of $\tilde{H}$ in H_j. Now, $B_j\tilde{H} = \tilde{H}$ and $B_j\tilde{H}^\perp = B_j(\tilde{\tilde{H}} \oplus H_j^\perp) = \tilde{\tilde{H}}$ and these spaces are complementary.

In the other case, when $\tilde{H} \subset H_j^\perp$ then $H_j \subset \tilde{H}^\perp$ so again we have that $B_j\tilde{H} = \{0\}$ and $B_j\tilde{H}^\perp$ are complementary. This completes the proof of the first part of the lemma.

To see the second part, let us first show that we get the decomposition

$$V = (V \cap H_j) \oplus (V \cap H_j^\perp). \tag{4.13}$$

This follows from the fact from [7] that since the datum is geometric then $(V, V^\perp)$ is a critical pair. Also from there [the proof of Lemma 7.12] we have that

$$\mathrm{tr}_H(B_jP_V) = \dim(B_jV)$$

where P_V is the orthogonal projection onto V and since B_jP_V is a contraction we get that there are independent vectors $\{v_l | l = 1, \ldots, \dim(B_jV)\}$ such that $\|B_jP_Vv_l\| = \|P_Vv_l\| = \|v_l\|$. The latter of these equalities says that $v_l \in V$ and then the former says that $v_l \in H_j$. Now for any vector $\tilde{v} \in V$ which is orthogonal to each v_l we must have that $B_jP_V\tilde{v}$ is the zero vector so $P_V\tilde{v} = \tilde{v}$ is in $H_j^\perp$. This proves that the decomposition (4.13) holds.

Let us now assume that $\tilde{H} = \cap_j H_j^a$ and show that

$$V = \left(V \cap \tilde{H}\right) \bigoplus \left(\sum_{j=1}^{m} (V \cap H_j^{a\perp})\right).$$

We see immediately that the space on the right hand side is a subspace of V and that each constituent of the sum is orthogonal to $\tilde{H}$ so the sum is a subspace of $V \cap \tilde{H}^\perp$. What is left to show is that the right hand side contains the whole of V. So, take a vector $u \in V$ which is orthogonal to each term in the sum on the right. It therefore lies in the orthogonal complement of $V \cap H_j^{a\perp}$ for each j. But from (4.13) we then get that it is in $V \cap H_j^a$ for each j and thus in $\cap_j(V \cap H_j^a)$ which equals $V \cap \tilde{H}$. To complete the proof of the lemma we note that the criticality of $V \cap \tilde{H}$ follows directly from Lemma 2.9. Then $(V \cap \tilde{H})^\perp$ is critical since the orthogonal complement of a critical space is critical in the geometric set-up and so again by Lemma 2.9 we get that $V \cap \tilde{H}^\perp = (V \cap \tilde{H})^\perp \cap V$ is critical. □

Remark 4.4. This lemma shows that the independent decomposition of H is also a critical decomposition and that any maximal critical decomposition of H is a refinement of the independent one.

We also note that a maximal critical decomposition is not unique. For example for Hölder's inequality, we can take any orthogonal basis $\{u_1, \ldots, u_n\}$ of H and $H = \oplus\langle u_i\rangle$ is a maximal critical decomposition. The independent decomposition here is simply H.

However, some parts of the decomposition are shared between any maximal critical decomposition. For example, in the Loomis–Whitney type situation when we let $\{e_i | i = 1, \ldots, 6\}$ be an orthogonal basis for H and B_j for $j = 1, 2, 3$ be the projection onto the span of e_{2j-1} and e_{2j} then any maximal critical decomposition is a refinement of the decomposition

$$H = \langle e_1, e_2\rangle \oplus \langle e_3, e_4\rangle \oplus \langle e_5, e_6\rangle$$

and this decomposition is the independent decomposition of H.

Furthermore, even $\tilde{\tilde{H}}_2$ need not have a unique maximal critical decomposition. As an example take the case when $\{e_i | i = 1, \ldots, 4\}$ is a orthogonal basis for H and B_j for $j = 1, 2, 3$ are the orthogonal projections onto $\langle e_1, e_3\rangle$, $\langle e_2, e_3\rangle$ and $\langle e_1 + e_2, e_3 + e_4\rangle$ respectively. Then $\tilde{\tilde{H}}_2 = H$ and

$$H = \langle e_1, e_2\rangle \oplus \langle e_3, e_4\rangle = \langle e_1 + e_3, e_2 + e_4\rangle \oplus \langle e_1 - e_3, e_2 - e_4\rangle$$

are two maximal critical decompositions of H.

We have from Lemma 1.16 that since the datum we are working with is geometric then for any critical subspace V, the orthogonal complement $V^\perp$ is also critical. Therefore, from the decomposition (4.13) we get for any j that

$$H = (V \cap H_j) \oplus (V \cap H_j^\perp) \oplus (V^\perp \cap H_j) \oplus (V^\perp \cap H_j^\perp). \tag{4.14}$$

This shows that $B_j^* B_j P_V = P_V B_j^* B_j$ and furthermore that

$$B_j P_V = B_j P_V B_j^* B_j. \tag{4.15}$$

With this set-up in hand we can state our main theorem as follows.

Theorem 4.5. *Let $((B_j), (p_j))$ be a geometric Brascamp–Lieb datum as above and let $\oplus_k \tilde{H}_k \oplus \tilde{\tilde{H}}_2$ be the independent decomposition of H.*

Assume that $(\tilde{f}_j)$ is an extremiser for this datum.

Then there exist integrable functions $G_k : H_k \to \mathbb{R}$, $k = 1, \ldots, k_0$, a critical decomposition $\tilde{H}_{k_0+1} \oplus \cdots \oplus \tilde{H}_{k_1}$ of $\tilde{\tilde{H}}_2$, positive constants c_j for $j = 1, \ldots, m$ and d_k for $k = k_0 + 1, \ldots, k_1$ and an element b from $\tilde{\tilde{H}}_2$ such that

$$\tilde{f}_j(x) = c_j \prod_{k=1}^{k_0} G_k(P_{j,k} B_j^* x) \prod_{k=k_0+1}^{k_1} e^{-\langle d_k P_{j,k} B_j^* x, P_{j,k}(B_j^* x + b)\rangle} \tag{4.16}$$

where $P_{j,k}$ is the orthogonal projection from H to $H_j \cap \tilde{H}_k$.

Conversely, all functions of this form are optimisers for this problem.

Proof. Since the datum $((B_j),(p_j))$ is extremisable it is a theorem from [7] that it is also gaussian-extremisable. Then the whole theory from the previous section applies and since we know that we have equality in (4.12) for gaussian optimisers we have for $(\tilde{f}_j)$ that

$$\frac{\partial}{\partial t}\int F(x,t)\,\mathrm{d}x = 0$$

for all $t > 0$ where we have continued with the notation introduced in the previous section. Let us fix a time $t > 0$ and suppress the dependence of the various quantities on it for the time being.

As the datum is geometric we get that M and S_j are the identity transformations and we may assume that the image of B_j is a subspace H_j of H and that $B_j^* : H_j \to H$ is the inclusion map. As noted above, the fact that $(\tilde{f}_j)$ is an extremiser means that the quantity (4.10) must be 0 for all $t > 0$ and all $x \in H$. We must then have $\langle (I-P)K^*, K^*\rangle$. We have that $\langle P\alpha, \alpha\rangle \leq \langle \alpha, \alpha\rangle$ for all vectors α and there is equality here if and only if $P\alpha = \alpha$, which means that α is in the image of P. We have that $P = TT^*$ where, since the Brascamp–Lieb datum is geometric,

$$T = \begin{pmatrix} --- & p_j^{\frac{1}{2}} B_1 & --- \\ & \vdots & \\ --- & p_j^{\frac{1}{2}} B_m & --- \end{pmatrix}.$$

We note that T is a linear transformation from H to $\oplus_j H_j$ and P is the projection onto the span of the column vectors of T. Therefore we see that the quantity (4.10) is 0 if and only if there exists a map $\beta : H \to H$ such that

$$K^*(x) = T\beta(x) \tag{4.17}$$

for almost every $x \in H$.

If we read off the rows in the above equation we find that $p_j^{-\frac{1}{2}}\nabla_{B_j x} k_j p_j = p_j^{\frac{1}{2}} B_j \beta(x)$ or

$$\nabla_{B_j x} k_j = B_j \beta(x) \tag{4.18}$$

since $Dk_j^* = \nabla k_j$. Then we see that

$$\nabla \log F(x) = \sum_{j=1}^{m} p_j B_j^* \nabla k_j(B_j x) = \sum_{j=1}^{m} p_j B_j^* B_j \beta(x) = \beta(x)$$

where we have used the geometricity of the datum for the last equality. We note from this equation that β is smooth and (4.17) must hold for all x.

By using (4.18) we can make the following calculation:

$$\begin{aligned} D_x(\log F)(b_j) &= \langle \nabla \log F(x), b_j\rangle = \langle \beta(x), b_j\rangle \\ &= \langle B_j\beta(x), e_j\rangle = \langle \nabla k_j(B_j x), e_j\rangle \end{aligned}$$

for any $b_j = B_j^* e_j$ where $e_j \in H_j$. If we differentiate this equality with respect to a vector $b_j^\perp = B_j^{\perp *} e_j^\perp$ where $e_j^\perp \in H_j^\perp$ and $B_j^\perp$ is the orthogonal projection onto $H_j^\perp$ we get that

$$D^2(\log F)(b_j, b_j^\perp) = 0$$

as the quantity on the far right hand side of the last chain of equalities is constant in the direction of $b_j^\perp$. This means that $\log F$ has the form $\log F = G_j^{\|}(B_j x) + G_j^\perp(B_j^\perp x)$ where $G_j^{\|}$ and $G_j^\perp$ are smooth.

Since we can make this calculation for any j we have established the equalities

$$\log F = G_j^{\|}(B_j x) + G_j^\perp(B_j^\perp x) = G_{j'}^{\|}(B_{j'} x) + G_{j'}^\perp(B_{j'}^\perp x) \tag{4.19}$$

for all j, j'.

In the following two lemmas we use this equality to determine the optimisers.

Lemma 4.6. *We can write*

$$\log F = \left(\sum_{k=1}^{k_0} G_{\tilde{H}_k}(P_{\tilde{H}_k} x) \right) + G_{\tilde{\tilde{H}}_2}(P_{\tilde{\tilde{H}}_2} x) \tag{4.20}$$

where

$$H = \left(\bigoplus_{k=1}^{k_0} \tilde{H}_k \right) \bigoplus \tilde{\tilde{H}}_2$$

is the independent decomposition of H.

Proof. The lemma will follow if we show that the second derivative of $\log F$ with respect to any pair of vectors from different components of this decomposition is identically zero. If the vectors come from two distinct independent subspaces, $\tilde{H}_k$ and $\tilde{H}_{k'}$ say, this follows immediately as there must be a j such that $\tilde{H}_k \subset H_j$ and $\tilde{H}_{k'} \subset H_j^\perp$ or the other way around. Then from (4.19) we can write

$$\log F = G_j^{\|}(B_j x) + G_j^\perp(B_j^\perp x)$$

and each differentiation will kill one component of the right hand side.

For the case when one of the vectors comes from $\tilde{\tilde{H}}_2$ we argue by contradiction. Let us assume that there is a vector b_k in some $\tilde{H}_k$ such that condition (4.19) does not guarantee that $\partial_{b_k} \partial_{b_0} \log F = 0$ for all $b_0 \in \tilde{\tilde{H}}_2$. Then there is a proper subspace U of $\tilde{\tilde{H}}_2$ such that $H_j^{a\perp} \cap \tilde{\tilde{H}}_2 \subset U$ for all j, where as before H_j^a is either H_j or $H_j^\perp$, such that $\tilde{H}_k = \cap_j H_j^a$. But then if we let $U^\perp$ be the orthogonal complement of U in $\tilde{\tilde{H}}_2$ we see from the orthogonality of each constituent of the maximal independent decomposition that $U^\perp \subset H_j^a$ for all j and so $U^\perp \subset \tilde{H}_k$ which is a contradiction to the fact that $U^\perp \subset \tilde{\tilde{H}}_2$. □

For the non-independent part we have the following lemma.

Lemma 4.7. *Assume that H has no independent subspaces. Then any optimiser is a gaussian.*

Proof. Let us take the gradient of (4.19). This gives

$$\nabla \log F = B_j^* \nabla G_j^{\|}(B_j x) + B_j^{\perp *} \nabla G_j^{\perp}(B_j^{\perp} x) = B_{j'}^* \nabla G_{j'}^{\|}(B_{j'} x) + B_{j'}^{\perp *} \nabla G_{j'}^{\perp}(B_{j'}^{\perp} x). \tag{4.21}$$

We want to use the Fourier transform to retrieve infromation from this equation. We first show that $\nabla \log F(x)$ has at most linear growth in x. Any function of polynomial growth defines a linear functional on the space of Schwartz functions and can thus be viewed as a tempered distribution and has a Fourier transform in that sense.

We are not able to establish the desired growth estimate for $\log F$ directly. Instead we replace the optimiser $(\tilde{f}_j)$ by $(\tilde{f}_j * f_{j0}) f_{j0}$ where $f_{j0}(x) = e^{-\|x\|_{H_j}^2}$. Since the datum we are working with is geometric then (f_{j0}) is an optimiser and $((\tilde{f}_j * f_{j0}) f_{j0})$ is an optimiser as well, see Lemma 6.3 in [7], and each function in this new tuple is a positive Schwartz function and remains a Schwartz function for all $t > 0$.

We will thus for the time being assume that the optimiser we are working with is a tuple of Schwartz functions. Now

$$\nabla \log F(x) = \sum_{j=1}^{m} p_j \frac{B_j^* \nabla \tilde{f}_j(B_j x)}{\tilde{f}_j(B_j x)}$$

so the linear growth of $\nabla \log F$ will follow if we show that $\frac{\nabla f_j(x_j)}{f_j(x_j)}$ has linear growth in $x_j = B_j x$. From (4.3) we see that

$$\frac{\nabla \tilde{f}_j(x_j)}{\tilde{f}_j(x_j)} = -\frac{1}{2t} \frac{\int_{H_j} (x_j - y_1) e^{-\frac{1}{4t}\|x_j - y_1\|^2} \tilde{f}_j(y_1)\, \mathrm{d}y_1}{\int_{H_j} e^{-\frac{1}{4t}\|x_j - y_1\|^2} \tilde{f}_j(y_1)\, \mathrm{d}y_1} \tag{4.22}$$

since in the geometric case $\tilde{A}_j = \mathrm{id}_{H_j}$.

Since $\tilde{f}$ is a positive Schwartz function we can find a C such that

$$\int_{\|y_1\| > C} \|y_1\| f_j(y_1)\, \mathrm{d}y_1 \leq \int_{\|y_1\| < C} f_j(y_1)\, \mathrm{d}y_1. \tag{4.23}$$

We may assume that $\|x_j\| > C$. We split the integral in the numerator of (4.22) in two parts according to whether $\|x_j - y_1\| \leq 2\|x_j\|$ or $\|x_j - y_1\| > 2\|x_j\|$. The first integral we can estimate by

$$2\|x_j\| \int_{H_j} e^{-\frac{1}{4t}\|x_j - y_1\|^2} \tilde{f}_j(y_1)\, \mathrm{d}y_1$$

and the contribution to the fraction from this term is therefore linear in $\|x_j\|$.

For the second part we note that if $\|x_j - y_1\| > 2\|x_j\|$ then $\|x_j - y_1\| < 2\|y_1\|$ and since $\|x_j\| > C$ we also get that $\|y_1\| > C$ and that $\|x_j - y_1\| > \|x_j - a\|$ for any point a such that $\|a\| < C$. We now see that

$$\begin{aligned} &\left\| \int_{\|x_j - y_1\| > 2\|x_j\|} (x_j - y_1) e^{-\frac{1}{4t}\|x_j - y_1\|^2} \tilde{f}_j(y_1)\, \mathrm{d}y_1 \right\| \\ &\leq 2 \int_{\|x_j - y_1\| > 2\|x_j\|} \|y_1\| e^{-\frac{1}{4t}\|x_j - y_1\|^2} \tilde{f}_j(y_1)\, \mathrm{d}y_1 \end{aligned}$$

and since the factor $\|x_j - y_1\|^2$ in the exponent here is larger for any point in the set $\{y_1 : \|x_j - y_1\| > 2\|x_j\|\}$ than for any point in the set $\{y_1 : \|y_1\| \leq C\}$ and the set $\{y_1 : \|x_j - y_1\| > 2\|x_j\|\}$ is contained in the set $\{y_1 : \|y_1\| > C\}$ we get by (4.23) that we can estimate the second integral by

$$\int_{\|y_1\| \leq C} e^{-\frac{1}{4t}\|x_j - y_1\|^2} \tilde{f}_j(y_1)\, \mathrm{d}y_1.$$

This establishes that $\nabla \log F$ has at most linear growth in x.

From (4.21) it is now evident that $B_j^* \nabla G_j^{\|}(B_j x)$ and $B_j^{\perp *} \nabla G_j^{\perp}(B_j^{\perp} x)$ have at most linear growth in x and we are therefore justified to take the Fourier transform of (4.21).

Let us now note that the Fourier transform of a function w of the form $w(x) = u(B_j x)$ is supported in $B_j^* H_j$. To see this, we take a test function ϕ and calculate as follows.

$$\begin{aligned} \int_H \hat{w}(x)\phi(x)\, \mathrm{d}x &= \int_H \int_H u(B_j x)\phi(\xi) e^{-i\langle x, \xi\rangle}\, \mathrm{d}\xi\, \mathrm{d}x \\ &= \int_{H_j} \int_{H_j^{\perp}} \int_{H_j} \int_{H_j^{\perp}} u(x_1)\phi(\xi_1, \xi_2) \\ &\qquad\qquad e^{-i\langle x_1, \xi_1\rangle} e^{-i\langle x_2, \xi_2\rangle}\, \mathrm{d}\xi_2\, \mathrm{d}\xi_1\, \mathrm{d}x_2\, \mathrm{d}x_1 \\ &= \int_{H_j} \int_{H_j} u(x_1)\phi(\xi_1, 0) e^{-i\langle x_1, \xi_1\rangle}\, \mathrm{d}\xi_1\, \mathrm{d}x_1 \end{aligned}$$

since

$$\int_{H_j^{\perp}} \int_{H_j^{\perp}} \phi(\xi_1, \xi_2) e^{-i\langle x_2, \xi_2\rangle}\, \mathrm{d}\xi_2\, \mathrm{d}x_2 = \phi(\xi_1, 0).$$

The last integral in the above expression is clearly 0 if ϕ is supported away from H_j. Now, equation (4.21) says that $\nabla \log F$ is the sum of a function which depends only on $B_j x$ and another which depends only on $B_j^{\perp} x$. It is therefore clear that the Fourier transform of $\nabla \log F$ is supported on $H_j \cup H_j^{\perp}$. Since this holds for any j we get that it is in fact supported on

$$\bigcap_j (H_j \cup H_j^{\perp}).$$

This intersection contains only the origin by the assumption that H has no independent subspaces.

It is well known that the Fourier transform of a distribution supported at the origin is a polynomial and we have thus established that $\nabla \log F$ is a polynomial and in fact a linear polynomial by the growth estimate above. Equation (4.18), together with the fact that $\nabla \log F = \beta$, gives that $\tilde{f}_j$ is a gaussian. Finally, we recall that we were working not with $\tilde{f}_j$ directly, but with $(\tilde{f}_j * f_{j0})f_{j0}$. However, if this function is a gaussian then since f_{j0} is a gaussian, it is clear that $\tilde{f}_j * f_{j0}$ is a gaussian and then by looking at the Fourier transform of this function we see that $\tilde{f}_j$ itself is a gaussian. $\square$

By the theory established in [7] for gaussian optimisers we know that since the datum is geometric there exists a maximal critical decomposition

$$\tilde{\tilde{H}}_2 = \bigoplus_{k=k_0+1}^{k_1} H_k$$

such that the purely quadratic term in the gaussian is the tensor product of multiples of the identity operator on each relevant H_k. Thus we have shown that with this decomposition we can write

$$\log F(x) = \sum_{k=1}^{k_0} G_{\tilde{H}_k}(P_{\tilde{H}_k}x) - \sum_{k=k_0+1}^{k_1} \langle d_k P_{\tilde{H}_k}x, P_{\tilde{H}_k}(x+b_k)\rangle$$

and so

$$\beta(x) = \nabla(\log F)(x) = \sum_{k=1}^{k_0} P^*_{\tilde{H}_k} \nabla G_{\tilde{H}_k}(P_{\tilde{H}_k}x) - \sum_{k=k_0+1}^{k_1} d_k P^*_{\tilde{H}_k} P_{\tilde{H}_k}(2x+b_k) \quad (4.24)$$

which gives

$$B_j\beta(x) = \sum_{k=1}^{k_0} B_j P^*_{\tilde{H}_k} \nabla G_{\tilde{H}_k}(P_{\tilde{H}_k}x) - \sum_{k=k_0+1}^{k_1} d_k B_j P^*_{\tilde{H}_k} P_{\tilde{H}_k}(2x+b_k).$$

Now, each term in the first part is zero unless $\tilde{H}_k \subset H_j$ in which case $P_{\tilde{H}_k} = P_{\tilde{H}_k} B_j^* B_j$. For the second part we have from (4.15) that $B_j P^*_{\tilde{H}_k} P_{\tilde{H}_k} = B_j P^*_{\tilde{H}_k} B_j^* B_j$. Therefore we have shown that $B_j\beta(x)$ depends only on $B_j x$ and moreover we see that $B_j\beta(x) = \nabla k_j(B_j x)$ where

$$k_j(x) = \sum_{\substack{k=1 \\ H_j \subset \tilde{H}_k}}^{k_0} G_{\tilde{H}_k}(P_{\tilde{H}_k}x) - \sum_{k=k_0+1}^{k_1} \langle d_k P_{H_j\cap\tilde{H}_k}x, P_{H_j\cap\tilde{H}_k}(x+b_k)\rangle. \quad (4.25)$$

This shows that $f_j(\cdot, t)$ must have the prescribed form for any $t > 0$. Since the set of tuples allowed by the theorem is a closed set in $L^1 \times \cdots \times L^1$ we get that

$(f_j) = (f_j(\cdot, 0))$ must also have this form as by the theory of the heat equation each f_j is the L^1 limit $\lim_{t\to 0} f_j(\cdot, t)$.

To prove the converse, that all tuples of the form (4.16) are optimisers, we first of all make the following remark. Assume that V is a critical subspace. Then from (4.14) we can write $H_j = (H_j \cap V) \oplus (H_j \cap V^\perp)$. Assume further that each f_j has the form $f_j(x) = f_{jV}(P_V B_j^* x) f_{jV^\perp}(P_{V^\perp} B_j^* x)$. Then

$$\int_H \prod_{j=1}^m f_j^{p_j}(B_j x)\,\mathrm{d}x = \int_V \prod_{j=1}^m f_{jV}^{p_j}(B_j|_V x_1)\,\mathrm{d}x_1 \int_{V^\perp} \prod_{j=1}^m f_{jV^\perp}^{p_j}(B_j|_{V^\perp} x_2)\,\mathrm{d}x_2$$

so (f_j) is an optimiser if (f_{jV}) and $(f_{jV^\perp})$ are optimisers for the data $((B_j|_V), (p_j))$ and $((B_j|_{V^\perp}), (p_j))$ respectively.

By repeating this splitting we may thus reduce to showing the following two things. Firstly, we must show that (f_j) with $f_j = c_j g$ where c_j is a constant and g is an integrable function is an optimiser in the case when $H_j = H$ for all j, $B_j = \mathrm{id}_H$, and the Brascamp–Lieb inequality reduces to

$$\int_H \prod_j f_j^{p_j}(x)\,\mathrm{d}x \le \prod_j \left(\int_H f_j(x)\,\mathrm{d}x\right)^{p_j}$$

with $\sum_j p_j = 1$. The proof is immediate by writing both sides of the inequality in terms of g.

Secondly, we must show that the gaussian tuple $(f_j) = (e^{-d\langle x, x + B_j b\rangle})$ with $d > 0$ and $b \in H$ is an optimiser for the Brascamp–Lieb inequality in the case when H has no independent subspaces and no proper critical subspace. However, even without these restrictions and only with the condition that $((B_j), (p_j))$ is geometric, it is well known that this tuple is an optimiser. $\square$

To conclude this chapter, let us drop the condition that $((B_j), (p_j))$ is geometric. However, from Theorem 7.13 in [7] we have that any extremisable datum is equivalent to a geometric datum. More specifically, equations (4.6) and (4.7) have a solution M and S_j with symmetric positive definite linear transformations and $((B'_j), (p_j))$ with $B'_j = S_j^{\frac{1}{2}} B_j M^{-\frac{1}{2}}$ is a geometric datum. Also, if (f_j) is an optimiser for $((B_j), (p_j))$ then $(f_j \circ S_j^{-\frac{1}{2}})$ is an optimiser for $((B'_j), (p_j))$ and conversely, if (f'_j) is an optimiser for $((B'_j), (p_j))$ then $(f'_j \circ S_j^{\frac{1}{2}})$ is an optimiser for $((B_j), (p_j))$.

As a direct consequence of this and Theorem 4.5 we get the following.

Theorem 4.8. *Let $((B_j), (p_j))$ be an extremisable Brascamp–Lieb datum.*

Assume that $(\tilde{f}_j)$ is an extremiser for this datum.

Let $((B'_j), (p_j))$ be the geometric datum equivalent to $((B_j), (p_j))$ and let M and S_j be such that $B'_j = S_j^{\frac{1}{2}} B_j M^{-\frac{1}{2}}$. Furthermore, let $\oplus_k \tilde{H}_k \oplus \tilde{\tilde{H}}_2$ be the independent *decomposition of* H *corresponding to the datum* $((B'_j),(p_j))$. *Then there exist integrable functions* $G_k : H_k \to \mathbb{R}$, $k = 1, \ldots, k_0$, *a critical decomposition* $\tilde{H}_{k_0+1} \oplus \cdots \oplus \tilde{H}_{k_1}$ *of* $\tilde{\tilde{H}}_2$, *positive constants* c_j *for* $j = 1, \ldots, m$ *and* d_k *for* $k = k_0+1, \ldots, k_1$ *and an element* b *from* $\tilde{\tilde{H}}_2$ *such that*

$$f_j(x) = c_j \prod_{k=1}^{k_0} G_k(P_{j,k} B_j'^* S_j^{\frac{1}{2}} x) \prod_{k=k_0+1}^{k_1} e^{-\langle d_k P_{j,k} B_j'^* S_j^{\frac{1}{2}} x, P_{j,k}(B_j'^* S_j^{\frac{1}{2}} x + b)\rangle} \qquad (4.26)$$

where $P_{j,k}$ *is the orthogonal projection from* H *to* $H_j \cap \tilde{H}_k$.

Conversely, all functions of this form are optimisers for this problem.

Chapter 5

The Hessian of the optimal transport potential

In this chapter we prove a generalisation of a theorem of Caffarelli from [14] and indicate how it applies to the methods which Barthe has used in [3] to work with Brascamp–Lieb inequalities.

To set things up, suppose we have two positive measures, μ_f and μ_g, on $\mathbb{R}^n$ which are absolutely continuous with respect to the Lebesgue measure and whose density functions are f and g respectively. Suppose further that the measures have finite second order moments and equal mass.

Then it is a well-known theorem of Brenier, see [10], [11] and the book [31], which asserts that there exists a unique positive measure π on $\mathbb{R}^n \times \mathbb{R}^n$ which has marginals μ_f and μ_g such that π is a minimiser for

$$I[\pi] := \int_{\mathbb{R}^n \times \mathbb{R}^n} |x-y|^2 \,\mathrm{d}\pi(x,y)$$

over all measures on $\mathbb{R}^n \times \mathbb{R}^n$ with these marginals. Furthermore, π has the form $\pi = (\mathrm{Id} \times \nabla\phi)\sharp\mu_f$ where $\sharp$ denotes the push-forward and $\nabla\phi$ is a uniquely determined gradient of a convex function which pushes μ_f forward to μ_g, i.e. $\nabla\phi\sharp\mu_f = \mu_g$. We call ϕ the optimal transportation potential.

The theorem of Caffarelli that we will generalise is then the following:

Theorem 5.1. *Suppose f and g are of the form*

$$f = (\det B)^{-\frac{1}{2}} e^{-\pi\langle B^{-1}\cdot,\cdot\rangle} \qquad \text{and} \qquad g = Ce^{-\pi\langle B^{-1}\cdot,\cdot\rangle - H}$$

where B is a positive definite symmetric linear transformation, H is convex and C is chosen so that $\int \mathrm{d}\mu_g = 1$. Then the optimal transport potential ϕ satisfies

$$\mathrm{Hess}(\phi, x) \leq I$$

where I is the identity transformation.

Here and onwards, where appropriate, this inequality is to be understood in the sense of positive definite linear transformations, that is $A \leq G$ if and only if $G - A$ is positive semi-definite.

We will prove the following generalisation:

Theorem 5.2. *Suppose f and g are of the form*

$$f = (\det(B^{-1}G))^{\frac{1}{2}} e^{-\pi\langle B^{-1}G\cdot,\cdot\rangle} * \mu \qquad \text{and} \qquad g = Ce^{-\pi\langle B^{-1}A^{-1}\cdot,\cdot\rangle - H}$$

where A, G and B are positive definite symmetric linear transformations, μ is a probability measure on $\mathbb{R}^n$, H is convex and C is chosen so that $\int g = 1$. Suppose also that $AB \leq GB = BG$. Then the optimal transport potential ϕ satisfies

$$\mathrm{Hess}(\phi, x) \leq G.$$

Note that we do not assume that A commutes with either B or G. Also note that it would be no restriction if we took the exponent in the definition of g to be $-\pi\langle B^{-1}G^{-1}\cdot,\cdot\rangle - H$ so the linear transformation A is superfluous.

To see this, note that

$$-\pi\langle B^{-1}A^{-1}\cdot,\cdot\rangle - H = -\pi\langle B^{-1}G^{-1}\cdot,\cdot\rangle - H'$$

where $H' = H + \pi\langle(B^{-1}A^{-1} - B^{-1}G^{-1})\cdot,\cdot\rangle$ and since it is known, see for example [22], p471, that the condition $AB \leq GB$ is equivalent to the condition $B^{-1}A^{-1} \geq B^{-1}G^{-1}$ we see that H' is convex if H is convex. However, we choose to include A in the definition of g because the case $g = \det A^{-\frac{1}{2}} e^{-\pi\langle A^{-1}\cdot,\cdot\rangle}$ will be important in Section 5.2.

The proof of Theorem 5.2 is the content of Section 5.1. It follows similar lines as the proof of Caffarelli in [14] but is somewhat more involved.

In Section 5.2 we use Theorem 5.2 to get results about Brascamp–Lieb and Reverse Brascamp–Lieb inequalities. We now summarise these results.

Let $B_j : \mathbb{R}^n \to \mathbb{R}^{n_j}$ be surjective linear transformations for $j = 1, \ldots, m$. Assume that $\cap_{j=1}^m \ker B_j = \{0\}$. Let us define the forms

$$J((f_j)_{j=1}^m) = \int_{\mathbb{R}^n} \prod_{j=1}^m f_j^{p_j}(B_j x)\,\mathrm{d}x$$

and

$$I((g_j)_{j=1}^m) = \int_{\mathbb{R}^n}^* \sup\left\{\prod_{j=1}^m g_j^{p_j}(y_j) : \sum_{j=1}^m p_j B_j^* y_j = x, y_j \in \mathbb{R}^n\right\}\mathrm{d}x.$$

We consider the inequalities

$$J((f_j)_{j=1}^m) \leq F \prod_{j=1}^{m} \left(\int f_j \right)^{p_j} \tag{5.1}$$

and

$$I((g_j)_{j=1}^m) \geq E \prod_{j=1}^{m} \left(\int g_j \right)^{p_j} \tag{5.2}$$

and ask what are the optimal values for E and F such that these inequalities hold for all non-negative integrable functions f_j and g_j. In [26], Lieb proved the fundamental result that (5.1) is exhausted by centred gaussians, meaning that the optimal value for F can be computed by considering only f_j of the form $e^{-\pi\langle A_j \cdot, \cdot \rangle}$ where A_j is a positive definite symmetric linear transformation. By using the well known fact that $\int e^{-\pi\langle A_j x, x\rangle}\, \mathrm{d}x = (\det A_j)^{-\frac{1}{2}}$ to calculate the integrals in (5.1) we thus get that the best constant is $F = D^{-\frac{1}{2}}$ where

$$D = \inf_{A_j} \left\{ \frac{\det(\sum_{j=1}^m p_j B_j^* A_j B_j)}{\prod_{j=1}^m (\det A_j)^{p_j}} \right\}.$$

In [3], Barthe used methods from the theory of optimal transportation to reprove Lieb's result and also at the same time prove the dual result that (5.2) is also exhausted by centred gaussians and that the best constant there is $E = D^{\frac{1}{2}}$.

We wish to extend the results of Barthe to the setting of generalised Brascamp–Lieb inequalities as introduced in Section 8 of [7]. We begin with the following definition:

Definition 5.3. Suppose G is a positive definite symmetric linear transformation and f and g are non-negative functions. We say that

1. f is of *class* G if f is the convolution of the gaussian $e^{-\pi\langle G \cdot, \cdot \rangle}$ with a positive measure and

2. g is of *inverse class* G if g has the form

$$g = e^{-\pi\langle G^{-1} \cdot, \cdot \rangle - H}$$

where H is a convex function.

These classes complement each other in the strategy of Barthe as will become clear in Section 5.2.

We now wish to consider inequalities (5.1) and (5.2) when f_j are functions of class G_j and g_j are of inverse class G_j. In this case, the inequalities are also exhausted by centred gaussians, restricted to the relevant class.

Specifically, we have the following theorem:

Theorem 5.4.

1. *(Generalised Brascamp–Lieb)*

$$J((f_j)_{j=1}^m) \le \frac{1}{\sqrt{D_{\mathbf{G}}}} \prod_{j=1}^m \left(\int f_j \right)^{p_j}$$

for all f_j of class G_j and

2. *(Generalised Reverse Brascamp–Lieb)*

$$I((g_j)_{j=1}^m) \ge \sqrt{D_{\mathbf{G}}} \prod_{j=1}^m \left(\int g_j \right)^{p_j}$$

for all g_j of inverse class G_j

where

$$D_{\mathbf{G}} = \inf_{A_j \le G_j} \left\{ \frac{\det(\sum_{j=1}^m p_j B_j^* A_j B_j)}{\prod_{j=1}^m (\det A_j)^{p_j}} \right\}.$$

Remark 5.5. The first part of the theorem has already been seen in [7] but the second part is new. In [7] it is also noted that we have $D_{\mathbf{G}} > 0$ if

$$\sum_{j=1}^m p_j B_j^* G_j B_j \ge B_l^* G_l B_l \tag{5.3}$$

for all $l = 1, \dots, m$ and in that case we have that $f_j(x) = e^{-\pi\langle G_j x, x\rangle}$ and $g_j(x) = e^{-\pi\langle G_j^{-1} x, x\rangle}$ are extremisers for (5.1) and (5.2) respectively so

$$D_{\mathbf{G}} = \frac{\det(\sum_{j=1}^m p_j B_j^* G_j B_j)}{\prod_{j=1}^m (\det G_j)^{p_j}}.$$

The proof of Theorem 5.4, which is in Section 5.2, follows the same steps as Barthe does in [3] but the added ingredient is Theorem 5.2.

We thank Robert McCann and Eric Carlen for bringing the paper [14] to our attention.

5.1 Proof of Theorem 5.2

First of all, let us note that it is straightforward to verify that the hypotheses of Brenier's Theorem are satisfied so that the potential ϕ is indeed well-defined.

We wish to prove that $\mathrm{Hess}(\phi, x) \le G$. However, for technical reasons which will become clear, we will make a couple of modifications. First of all, we replace g by $g^r : \bar{B}_r \mapsto \mathbb{R}$ given by $g^r(x) = C' g(x)$ for $|x| \le r$ where $\bar{B}_r$ denotes the closed ball in $\mathbb{R}^n$ of radius r and C' is a normalising constant chosen so that $\int g^r = 1$.

We specify that the function which transports $f\,\mathrm{d}\mathcal{L}^n$ to $g^r\,\mathrm{d}\mathcal{L}^n|_{\bar{B}_r}$ is $\nabla\phi^r$ where ϕ^r is convex. Secondly, we replace the Hessian by a finite difference quotient

$$\frac{\phi^r(x+h\alpha)+\phi^r(x-h\alpha)-2\phi^r(x)}{h^2}$$

for some fixed $h \geq 0$.

We are therefore interested in the function

$$K(x,\alpha) := \langle G\alpha,\alpha\rangle - \frac{\phi^r(x+h\alpha)+\phi^r(x-h\alpha)-2\phi^r(x)}{h^2}$$

which is C^2 on $\mathbb{R}^n \times S^{n-1}$ and we wish to show that K is non-negative.

Our strategy will be to show that at any point where K has a local minimum then K is non-negative. From the convexity of ϕ^r it is clear that

$$K(x,\alpha) \leq \langle G\alpha,\alpha\rangle \tag{5.4}$$

for any α in S^{n-1}. So if we can show that in the limit as x tends to infinity this inequality becomes an equality then it is guaranteed that K has a local minimum which is also a global minimum.

If we work with g and ϕ directly we cannot hope that (5.4) becomes an equality in the limit as can be easily seen from the example

$$f = (\det G)^{\frac{1}{2}} e^{-\pi\langle G\cdot,\cdot\rangle} \qquad g = (\det A)^{-\frac{1}{2}} e^{-\pi\langle A^{-1}\cdot,\cdot\rangle}$$

where it is easy to confirm that the transport map is given by

$$\nabla\phi(x) = A^{\frac{1}{2}}G^{\frac{1}{2}}x$$

so $\mathrm{Hess}(\phi,x) = A^{\frac{1}{2}}G^{\frac{1}{2}}$ for all x and if $A < G$ this does not equal G. So the reason why we use g^r instead of g is that the resulting ϕ^r has much better behaviour at infinity than we can expect of ϕ as we shall see in Lemma 5.6 at the end of this section and in the discussion thereafter.

However, we have only been able to prove this good behaviour of ϕ^r at infinity for the finite difference quotient and not the Hessian itself. That is one reason why we use the finite difference but another is that if we use the finite difference then we only need to know that ϕ^r is C^2 and this follows directly from Caffarelli's regularity theory because H is convex and therefore locally Lipschitz continuous, see Chapter 4 of [31].

Then we get the pointwise Monge-Ampère equation which is the key equation relating ϕ^r to f and g^r;

$$g^r(\nabla\phi^r(x))\det(\mathrm{Hess}(\phi^r,x)) = f(x). \tag{5.5}$$

Let us now assume that K has a local minimum at (x_0, α_0). We have that

$$\begin{aligned}\partial_{x_i} K(x_0,\alpha_0) =& \partial_{x_i}\left(-\frac{\phi^r(x_0+h\alpha_0)+\phi^r(x_0-h\alpha_0)-2\phi^r(x_0)}{h^2}\right)\\ =& -\frac{\phi_i^r(x_0+h\alpha_0)+\phi_i^r(x_0-h\alpha_0)-2\phi_i^r(x_0)}{h^2}\end{aligned}$$

and since (x_0, α_0) is a local minimum we get that

$$\phi_i^r(x_0+h\alpha_0)+\phi_i^r(x_0-h\alpha_0)-2\phi_i^r(x_0)=0.$$

Since this holds for $i = 1, \ldots, n$ we get that

$$\nabla\phi^r(x_0+h\alpha_0)+\nabla\phi^r(x_0-h\alpha_0)-2\nabla\phi^r(x_0)=0. \tag{5.6}$$

Also, we calculate

$$\partial_{\alpha^\perp} K(x_0,\alpha_0) = 2G\alpha_0\cdot\alpha^\perp - \frac{\nabla\phi^r(x_0+h\alpha_0)\cdot h\alpha^\perp - \nabla\phi^r(x_0-h\alpha_0)\cdot h\alpha^\perp}{h^2}$$

where $\partial_{\alpha^\perp}$ denotes the directional derivative in the direction of $\alpha^\perp$. Since (x_0, α_0) is a local minimum we get that

$$\left(2G\alpha_0 - \frac{\nabla\phi^r(x_0+h\alpha_0)-\nabla\phi^r(x_0-h\alpha_0)}{h}\right)\cdot\alpha^\perp = 0$$

for any unit vector $\alpha^\perp$ which is perpendicular to α_0. We can interpret this as saying that there exists a $\lambda \in \mathbb{R}$ such that

$$G\alpha_0 = \frac{\nabla\phi^r(x_0+h\alpha_0)-\nabla\phi^r(x_0-h\alpha_0)}{2h} - \lambda\alpha_0. \tag{5.7}$$

Solving this equation together with (5.6) gives

$$\nabla\phi^r(x_0\pm h\alpha_0) = \nabla\phi^r(x_0)\pm h(G\alpha_0+\lambda\alpha_0). \tag{5.8}$$

Let us now take the relevant finite difference in (5.5). This gives

$$\begin{aligned}&\log\det(\mathrm{Hess}(\phi^r,x_0+h\alpha_0))+\log\det(\mathrm{Hess}(\phi^r,x_0-h\alpha_0))\\ &\quad -2\log\det(\mathrm{Hess}(\phi^r,x_0))\\ &=\log f(x_0+h\alpha_0)+\log f(x_0-h\alpha_0)-2\log f(x_0)\\ &\quad -\left[\log g^r(\nabla\phi^r(x_0+h\alpha_0))+\log g^r(\nabla\phi^r(x_0-h\alpha_0))\right.\\ &\qquad \left.-2\log g^r(\nabla\phi^r(x_0))\right].\end{aligned} \tag{5.9}$$

Dealing with the individual terms of this equation will be our main task in what follows. We know that log det is a concave function so the graph of the tangent plane at any point lies above the graph of the function. If we use this at $\mathrm{Hess}(\phi^r, x_0)$ we get that the left hand side of the equation is less than

$$\mathrm{D}(\log\det)(\mathrm{Hess}(\phi^r,x_0))\cdot E$$

where

$$E := \operatorname{Hess}(\phi^r, x_0 + h\alpha_0) + \operatorname{Hess}(\phi^r, x_0 - h\alpha_0) - 2\operatorname{Hess}(\phi^r, x_0)$$

and $\mathrm{D}(\log\det)$ is the total derivative of $\log\det$. It is clear that E is the Hessian of the function

$$x \mapsto \phi^r(x + h\alpha_0) + \phi^r(x - h\alpha_0) - 2\phi^r(x)$$

which by our assumptions attains a maximum at x_0 and therefore by the second derivative test we see that E is negative semi-definite. By expanding the determinant by minors and using Cramer's formula we can see that

$$\mathrm{D}(\log\det)(\operatorname{Hess}(\phi^r, x_0)) \cdot E = \sum_{i,j}((\operatorname{Hess}(\phi^r, x_0))^{-1})_{ij}E_{ij}$$

Now, for any two positive semi-definite symmetric matrices E and C we can write $\sum_{i,j} C_{ij}E_{ij}$ as $\operatorname{tr}(EC)$ and if $E^{\frac{1}{2}}$ and $C^{\frac{1}{2}}$ are the positive semi-definite symmetric square roots of E and C respectively we can calculate

$$\operatorname{tr}(EC) = \operatorname{tr}(E^{\frac{1}{2}}E^{\frac{1}{2}}C^{\frac{1}{2}}C^{\frac{1}{2}}) = \operatorname{tr}(E^{\frac{1}{2}}C^{\frac{1}{2}}C^{\frac{1}{2}}E^{\frac{1}{2}}) = \operatorname{tr}(E^{\frac{1}{2}}C^{\frac{1}{2}}(E^{\frac{1}{2}}C^{\frac{1}{2}})^T) \geq 0.$$

This tells us that the term we are working on is non-positive because E is negative semi-definite and $(\operatorname{Hess}(\phi^r, x_0))^{-1}$ is positive definite.

Let us then examine the right hand side of (5.9). The second directional derivative of $\log(f)$ in the direction α_0 is given by

$$\frac{f_{\alpha_0\alpha_0}(x)}{f(x)} - \left(\frac{f_{\alpha_0}(x)}{f(x)}\right)^2.$$

We therefore perform the following calculation:

$$\begin{aligned} f(x) =& (\det(B^{-1}G))^{\frac{1}{2}} \int e^{-\pi\langle B^{-1}G(x-y),(x-y)\rangle}\,\mathrm{d}\mu(y), \\ f_{\alpha_0}(x) =& (\det(B^{-1}G))^{\frac{1}{2}} \int -2\pi\langle B^{-1}G\alpha_0, x-y\rangle e^{-\pi\langle B^{-1}G(x-y),(x-y)\rangle}\,\mathrm{d}\mu(y) \end{aligned}$$

and

$$\begin{aligned} f_{\alpha_0\alpha_0}(x) =& (\det(B^{-1}G))^{\frac{1}{2}} \int 4\pi^2(\langle B^{-1}G\alpha_0, x-y\rangle)^2 e^{-\pi\langle B^{-1}G(x-y),(x-y)\rangle}\,\mathrm{d}\mu(y) \\ &-(\det(B^{-1}G))^{\frac{1}{2}} \int 2\pi\langle B^{-1}G\alpha_0, \alpha_0\rangle e^{-\pi\langle B^{-1}G(x-y),(x-y)\rangle}\,\mathrm{d}\mu(y). \end{aligned}$$

The Cauchy-Schwarz inequality gives us that

$$\begin{aligned} (f_{\alpha_0}(x))^2 =& (\det(B^{-1}G)) \left(\int -2\pi\langle B^{-1}G\alpha_0, x-y\rangle e^{-\pi\langle B^{-1}G(x-y),(x-y)\rangle}\,\mathrm{d}\mu(y)\right)^2 \\ \leq& (\det(B^{-1}G))^{\frac{1}{2}} \int e^{-\pi\langle B^{-1}G(x-y),(x-y)\rangle}\,\mathrm{d}\mu(y) \\ &\cdot(\det(B^{-1}G))^{\frac{1}{2}} \int 4\pi^2(\langle B^{-1}G\alpha_0, x-y\rangle)^2 e^{-\pi\langle B^{-1}G(x-y),(x-y)\rangle}\,\mathrm{d}\mu(y) \end{aligned}$$

and this tells us that

$$\frac{f_{\alpha_0\alpha_0}(x)}{f(x)} - \left(\frac{f_{\alpha_0}(x)}{f(x)}\right)^2 \geq -2\pi\langle B^{-1}G\alpha_0, \alpha_0\rangle.$$

By integrating twice we can estimate the terms involving f from below by

$$-2\pi h^2\langle B^{-1}G\alpha_0, \alpha_0\rangle.$$

The terms in (5.9) involving g^r can be split into a sum of three parts. The first part comes from the normalising constant C and this will equal $2C - 2C = 0$. The third part will be

$$\begin{aligned} &H(\nabla\phi^r(x_0 + h\alpha_0)) + H(\nabla\phi^r(x_0 - h\alpha_0)) - 2H(\nabla\phi^r(x_0)) \\ &\geq \mathrm{D}H(\nabla\phi^r(x_0)) \cdot (\nabla\phi^r(x_0 + h\alpha_0) + \nabla\phi^r(x_0 - h\alpha_0) - 2\nabla\phi^r(x_0)) = 0 \end{aligned}$$

where we have used the convexity of H and the condition (5.6).

The second part is

$$\begin{aligned} &\pi\langle B^{-1}A^{-1}\nabla\phi^r(x_0 + h\alpha_0), \nabla\phi^r(x_0 + h\alpha_0)\rangle \\ &+\pi\langle B^{-1}A^{-1}\nabla\phi^r(x_0 - h\alpha_0), \nabla\phi^r(x_0 - h\alpha_0)\rangle - 2\pi\langle B^{-1}A^{-1}\nabla\phi^r(x_0), \nabla\phi^r(x_0)\rangle. \end{aligned}$$

Using (5.8) we can simplify this to

$$2\pi h^2\langle B^{-1}A^{-1}(G\alpha_0 + \lambda\alpha_0), G\alpha_0 + \lambda\alpha_0\rangle.$$

When we take all these calculations together we see that we have reduced (5.9) to the simple inequality

$$0 \geq -2\pi h^2\langle B^{-1}G\alpha_0, \alpha_0\rangle + 2\pi h^2\langle B^{-1}A^{-1}(G\alpha_0 + \lambda\alpha_0), G\alpha_0 + \lambda\alpha_0\rangle$$

which says

$$\langle B^{-1}G\alpha_0, \alpha_0\rangle \geq \langle B^{-1}A^{-1}(G\alpha_0 + \lambda\alpha_0), G\alpha_0 + \lambda\alpha_0\rangle.$$

Now we want to use the fact that $B^{-1}A^{-1} \geq B^{-1}G^{-1}$. As we have already mentioned this follows from the condition that $GB \geq AB$. Using this we see that

$$\langle B^{-1}G\alpha_0, \alpha_0\rangle \geq \langle B^{-1}G^{-1}(G\alpha_0 + \lambda\alpha_0), G\alpha_0 + \lambda\alpha_0\rangle$$

and by expanding the right hand side we get that

$$0 \geq 2\lambda\langle B^{-1}\alpha_0, \alpha_0\rangle + \lambda^2\langle B^{-1}G^{-1}\alpha_0, \alpha_0\rangle$$

where we have used the assumption that B and G commute. This is a quadratic expression in λ and since both coefficients are positive we can deduce from this that $\lambda \leq 0$. By taking the inner product of (5.7) with α_0 and using this we get that

$$\langle G\alpha_0, \alpha_0 \rangle \geq \frac{\nabla\phi^r(x_0 + h\alpha_0) \cdot \alpha_0 - \nabla\phi^r(x_0 - h\alpha_0) \cdot \alpha_0}{2h}. \tag{5.10}$$

Unfortunately, when we want to use this equation to tell us something about $K(x, \alpha)$ we are forced to take a less than optimal route. This is because we only have information about the behaviour of $\nabla\phi^r$ at $x_0 \pm h\alpha_0$ so the best we can do is to say that

$$\begin{aligned} &\frac{\phi^r(x_0 + h\alpha_0) + \phi^r(x_0 - h\alpha_0) - 2\phi^r(x_0)}{h^2} \\ &\leq \frac{\nabla\phi^r(x_0 + h\alpha_0) \cdot \alpha_0 - \nabla\phi^r(x_0 - h\alpha_0) \cdot \alpha_0}{h} \leq 2\langle G\alpha_0, \alpha_0 \rangle \end{aligned}$$

so we conclude that

$$K(x, \alpha) \geq -\langle G\alpha, \alpha \rangle$$

for all $x \in \mathbb{R}^n$ and $\alpha \in S^{n-1}$. In terms of ϕ^r this says that

$$\frac{\phi^r(x + h\alpha) + \phi^r(x - h\alpha) - 2\phi^r(x)}{h^2} \leq 2\langle G\alpha, \alpha \rangle. \tag{5.11}$$

This misses what we intended to prove by a factor of 2 but we can get around that by iterating. This is the same problem that is encountered in [14] but it is addressed in [15] and we follow that argument here.

Let us assume we have this estimate (5.11) with the factor of 2 replaced by a number a greater than 1, uniformly in h. We have until now assumed that h is a fixed positive number but if we temporarily allow it to pass to 0 we get that

$$\langle \mathrm{Hess}(\phi^r, x)\alpha, \alpha \rangle \leq a\langle G\alpha, \alpha \rangle.$$

Then we notice that

$$\begin{aligned} &\phi^r(x + h\alpha) + \phi^r(x - h\alpha) - 2\phi^r(x) \\ &\leq \int_0^h |\nabla\phi^r(x_0 + t\alpha_0) \cdot \alpha_0 - \nabla\phi^r(x_0 - t\alpha_0) \cdot \alpha_0| \, \mathrm{d}t \end{aligned}$$

and we have two ways of estimating the quantity within the absolute values. By looking at its derivative we get the upper bound $2t \cdot a\langle G\alpha, \alpha \rangle$ and we also still have the bound $2h \cdot \langle G\alpha, \alpha \rangle$ from (5.11). By replacing the integrand by the better of these we get

$$\frac{\phi^r(x + h\alpha) + \phi^r(x - h\alpha) - 2\phi^r(x)}{h^2} \leq \frac{2a - 1}{a}\langle G\alpha, \alpha \rangle.$$

Now, $a_1 = 2$, $a_{n+1} = \frac{2a_n - 1}{a_n}$ defines a decreasing sequence tending to 1 and by passing to this limit we get what we intended to prove.

Finally, since $g^r \to g$ in L^1 as $r \to \infty$ we get that

$$\frac{\phi(x + h\alpha) + \phi(x - h\alpha) - 2\phi(x)}{h^2} \leq \langle G\alpha, \alpha \rangle$$

and using that ϕ is twice continuously differentiable we can take the limit as $h \to 0$ and get

$$\langle \mathrm{Hess}(\phi, x)\alpha, \alpha \rangle \leq \langle G\alpha, \alpha \rangle$$

and this is what we intended to prove.

Let us now prove the lemma we left behind.

Lemma 5.6. *For any fixed r we have that*

$$\lim_{x\to\infty} \left| \nabla\phi^r(x) - r\frac{x}{|x|} \right| = 0.$$

Proof. Let us fix the vector $y = \nabla\phi^r(x)$ and look at the set

$$\Gamma_y := \{y' \in \mathbb{R}^n : \mathrm{angle}(x, y' - y) \leq \frac{\pi}{4}\}.$$

This is a cone originating from y, pointing in the direction of x. Since $\nabla\phi^r$ is an optimal transport plan, it is known from [20] that the support of its graph is cyclically monotone. This means that if we take $x_1, \ldots, x_m \in \mathbb{R}^n$ and let $y_i = \nabla\phi^r(x_i)$ then

$$\sum_{i=1}^{m} |x_i - y_i|^2 \leq \sum_{i=1}^{m} |x_i - y_{i-1}|^2 \tag{5.12}$$

where we let $y_0 = y_m$. In particular, if we take $y' = \nabla\phi^r(x')$ where $y' \in \Gamma_y$ and apply the inequality to x, x' and y, y' then we get that

$$\langle x' - x, y' - y \rangle \geq 0$$

and since $\mathrm{angle}(x, y' - y) \leq \frac{\pi}{4}$ we get that $\mathrm{angle}(x, x' - x) \leq \frac{3\pi}{4}$ so the preimage of Γ_y is contained in the concave cone

$$\Gamma_x := \{x' \in \mathbb{R}^n : \mathrm{angle}(x, x' - x) \leq \frac{3\pi}{4}\}.$$

The complement of the cone Γ_x contains a ball around the origin of radius $2^{-\frac{1}{2}}x$ and as x tends to infinity the mass of f on the complement of this ball, and thus on Γ_x, will tend to 0. We then see that

$$(\inf_{x\in B_r} g^r(x))|\Gamma_y \cap B_r| \leq (g^r \,\mathrm{d}\mathcal{L}^n)(\Gamma_y \cap B_r) = (g^r \,\mathrm{d}\mathcal{L}^n)(\Gamma_y) \leq (f \,\mathrm{d}\mathcal{L}^n)(\Gamma_x)$$

and by compactness we have that g^r is bounded away from 0 on B_r so we get that

$$\lim_{x\to\infty} |\Gamma_y \cap B_r| = 0$$

so by the geometry of the problem we get the desired conclusion. $\square$

With this lemma in hand we can crudely estimate the finite difference

$$\frac{\phi^r(x+h\alpha)+\phi^r(x-h\alpha)-2\phi^r(x)}{h^2}$$

by

$$\frac{\nabla\phi^r(x+h\alpha)\cdot\alpha-\nabla\phi^r(x-h\alpha)\cdot\alpha}{2h}$$

which tends to 0 as x tends to infinity. Therefore we get

$$\lim_{x\to\infty} K(x,\alpha)=\langle G\alpha,\alpha\rangle$$

and as already mentioned, this guarantees the existence of a global minimiser.

5.2 Proof of Theorem 5.4

Firstly, we note that if $D_{\mathbf{G}}=0$ the theorem has no content so in the following we will assume that $D_{\mathbf{G}}>0$. Let us define

$$E_{\mathbf{G}} := \inf\left\{\frac{I((f_j)_{j=1}^m)}{\prod_{j=1}^m\left(\int f_j\right)^{p_j}} : f_j \text{ of inverse class } G_j\right\}$$

and

$$F_{\mathbf{G}} := \sup\left\{\frac{J((f_j)_{j=1}^m)}{\prod_{j=1}^m\left(\int f_j\right)^{p_j}} : f_j \text{ of class } G_j\right\}.$$

Our aim is then to prove that

$$E_{\mathbf{G}}=\sqrt{D_{\mathbf{G}}} \qquad \text{and} \qquad F_{\mathbf{G}}=\frac{1}{\sqrt{D_{\mathbf{G}}}}.$$

To begin with, let us define

$$E_{\mathbf{G},g} := \inf\left\{\frac{I((f_j)_{j=1}^m)}{\prod_{j=1}^m\left(\int f_j\right)^{p_j}} : f_j \text{ centred gaussian of inverse class } G_j\right\}$$

and

$$F_{\mathbf{G},g} := \sup\left\{\frac{J((f_j)_{j=1}^m)}{\prod_{j=1}^m\left(\int f_j\right)^{p_j}} : f_j = e^{-\pi\langle A_j\cdot,\cdot\rangle} \text{ for } A_j\leq G_j\right\}.$$

Since the infimum for $E_{\mathbf{G},g}$ is taken over a smaller class of functions than the infimum for $E_{\mathbf{G}}$ we see that $E_{\mathbf{G},g}\geq E_{\mathbf{G}}$. Also, by calculating the convolution of $\det(G_j)^{\frac{1}{2}}e^{-\pi\langle G_j\cdot,\cdot\rangle}$ with the measure which has density function $\det(F_j)^{\frac{1}{2}}e^{-\pi\langle F_j\cdot,\cdot\rangle}$ where $F_j=G_j(G_j-A_j)^{-1}G_j-G_j$ we can see that if $A_j\leq G_j$ then $e^{-\pi\langle A_j\cdot,\cdot\rangle}$ is of class G_j so $F_{\mathbf{G}}\geq F_{\mathbf{G},g}$.

We now state three lemmas which will guide us through the proof.

Lemma 5.7.

$$F_{\mathbf{G},g} = \frac{1}{\sqrt{D_{\mathbf{G}}}}.$$

Lemma 5.8.

$$E_{\mathbf{G},g} F_{\mathbf{G},g} = 1.$$

Lemma 5.9. *If g_j is of inverse class G_j and f_j is of class G_j then*

$$I(g_1, \ldots, g_m) \geq D_{\mathbf{G}} J(f_1, \ldots, f_m).$$

To see how the result follows, note that by Lemma 5.9 we get that $E_{\mathbf{G}} \geq D_{\mathbf{G}} F_{\mathbf{G}}$ and from that we get the string of inequalities

$$\sqrt{D_{\mathbf{G}}} = E_{\mathbf{G},g} \geq E_{\mathbf{G}} \geq D_{\mathbf{G}} F_{\mathbf{G}} \geq D_{\mathbf{G}} F_{\mathbf{G},g} = \sqrt{D_{\mathbf{G}}}$$

so we get equality all the way and this gives the theorem.

All that remains is to give a proof of the three lemmas. Note first that centred gaussians of inverse class G_j are exactly those of the form $e^{-\pi\langle A_j^{-1}\cdot,\cdot\rangle}$ where $A_j \leq G_j$ so for the first lemma we take $f_j = e^{-\pi\langle A_j\cdot,\cdot\rangle}$. Then

$$J((f_j)_{j=1}^m) = \int_{\mathbb{R}^n} e^{-\pi \sum_{j=1}^m p_j \langle A_j B_j x, B_j x\rangle}\, \mathrm{d}x = \int_{\mathbb{R}^n} e^{-\pi Q(x)}\, \mathrm{d}x$$

where

$$Q(x) = \left\langle \sum_{j=1}^m p_j B_j^* A_j B_j x, x \right\rangle.$$

The fact that $\int e^{-\pi\langle Ax,x\rangle}\, \mathrm{d}x = (\det A)^{-\frac{1}{2}}$ for any positive definite linear transformation A gives

$$\frac{J((f_j)_{j=1}^m)}{\prod_{j=1}^m \left(\int f_j\right)^{p_j}} = \left(\frac{\prod_{j=1}^m (\det A_j)^{p_j}}{\det\left(\sum_{j=1}^m p_j B_j^* A_j B_j\right)} \right)^{\frac{1}{2}}$$

so $F_{\mathbf{G},g} = \frac{1}{\sqrt{D_{\mathbf{G}}}}$.

For the second lemma let $g_j = e^{-\pi\langle A_j^{-1}\cdot,\cdot\rangle}$. Then $I((g_j)_{j=1}^m) = \int e^{-\pi R(x)}\, \mathrm{d}x$ where

$$R(x) = \inf\left\{ \sum_{j=1}^m p_j \langle A_j^{-1} x_j, x_j\rangle : x = \sum_{j=1}^m p_j B_j^* x_j \text{ where } x_j \in \mathbb{R}^{n_j} \right\}.$$

We recall that the dual of a quadratic form Q is defined by

$$Q^*(x) = \sup\{|\langle x, y\rangle|^2 : Q(y) \leq 1\}.$$

It is shown in the proof of Lemma 2 in [3] that R is a quadratic form and that $R = Q^*$. Then we see that

$$\frac{I((g_j)_{j=1}^m)}{\prod_{j=1}^m \left(\int g_j\right)^{p_j}} = \left(\frac{\prod_{j=1}^m (\det A_j^{-1})^{p_j}}{\det(R)}\right)^{\frac{1}{2}}$$

and since $\det R = \det Q^* = (\det Q)^{-1}$ we get that $E_{\mathbf{G},g} F_{\mathbf{G},g} = 1$.

For the third lemma take f_j to be of class G_j and g_j to be of inverse class G_j. We may assume that $\int f_j = \int g_j = 1$. Then f_j and g_j satisfy the conditions of Theorem 5.2 so from that we get that there exists a C^2 transport potential ϕ_j such that

$$g_j(\nabla \phi_j(x)) \det(\mathrm{Hess}(\phi_j, x)) = f_j(x)$$

and $\mathrm{Hess}(\phi_j, x) \le G_j$ for all $x \in \mathbb{R}^{n_j}$. Define $\Theta(y) := \sum_{j=1}^m p_j B_j^* \nabla \phi_j(B_j y)$. Then the Jacobian of Θ at y is

$$\det\left(\sum_{j=1}^m p_j B_j^* \,\mathrm{Hess}(\phi_j, B_j y) B_j\right)$$

and this is positive by the assumption that $D_{\mathbf{G}} \ge 0$. We then repeat the calculations from the proof of Lemma 3 in [3]

$$\begin{aligned}
J((f_j)_{j=1}^m) &= \int_{\mathbb{R}^n} \prod_{j=1}^m f_j^{p_j}(B_j y)\,\mathrm{d}y \\
&= \int_{\mathbb{R}^n} \prod_{j=1}^m g_j^{p_j}(\nabla\phi_j(B_j y)) \prod_{j=1}^m (\det(\mathrm{Hess}(\phi_j, B_j y)))^{p_j}\,\mathrm{d}y \\
&\overset{(*)}{\le} \frac{1}{D_{\mathbf{G}}} \int_{\mathbb{R}^n} \prod_{j=1}^m g_j^{p_j}(\nabla\phi_j(B_j y)) \det\left(\sum_{j=1}^m p_j B_j^* \,\mathrm{Hess}(\phi_j, B_j y) B_j\right) \mathrm{d}y \\
&\le \frac{1}{D_{\mathbf{G}}} \int_{\mathbb{R}^n}^* \sup_{\substack{\Theta(y)= \\ \sum p_j B_j^* x_j}} \left(\prod_{j=1}^m g_j^{p_j}(x_j)\right) \det\left(\sum_{j=1}^m p_j B_j^* \,\mathrm{Hess}(\phi_j, B_j y) B_j\right) \mathrm{d}y
\end{aligned}$$

where the added ingredient is that in step $(*)$ we have used that $\mathrm{Hess}(\phi, x) \le G$ so that the inequality follows from the definition of $D_{\mathbf{G}}$. We can therefore make the change of variables $z = \Theta(y)$ and get

$$\begin{aligned}
J((f_j)_{j=1}^m) &\le \frac{1}{D_{\mathbf{G}}} \int_{\mathbb{R}^n}^* \sup_{z=\sum_{j=1}^m p_j B_j^* x_j} \left(\prod_{j=1}^m g_j^{p_j}(x_j)\right) \mathrm{d}z \\
&= I((g_j)_{j=1}^m).
\end{aligned}$$

This proves the third lemma.

Chapter 6

A multilinear generalisation of the Hilbert transform and fractional integration

6.1 A singular integral

We wish to study the n-linear form given formally by

$$\Lambda(f_1, \ldots, f_n) := \int_{(\mathbb{R}^{n-1})^n} \frac{f_1(x_1) \cdots f_n(x_n)}{\det \left(\begin{smallmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{smallmatrix} \right)} \, \mathrm{d}x_1 \, \mathrm{d}x_2 \ldots \mathrm{d}x_n \tag{6.1}$$

where $x_i \in \mathbb{R}^{n-1}$. For $n = 2$ we have

$$\Lambda(f, g) = \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{f(x)g(y)}{\det \left(\begin{smallmatrix} 1 & 1 \\ x & y \end{smallmatrix} \right)} \, \mathrm{d}x \, \mathrm{d}y = \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{f(x)g(y)}{y - x} \, \mathrm{d}x \, \mathrm{d}y = \pi \langle Hf, g \rangle$$

where Hf denotes the Hilbert transform of f so in this case Λ is the bilinear form associated to the Hilbert transform. For $n \geq 3$ we can see Λ as an n-linear generalisation of the Hilbert transform.

There is a closely related form defined for n functions on the unit sphere S^{n-1} given by

$$\Lambda_S(f_1, \ldots, f_n) := \int_{(S^{n-1})^n} \frac{f_1(\omega_1) \cdots f_n(\omega_n)}{\det(\omega_1, \ldots, \omega_n)} \, \mathrm{d}\omega_1 \, \mathrm{d}\omega_2 \ldots \mathrm{d}\omega_n. \tag{6.2}$$

The integrals (6.1) and (6.2) are not absolutely convergent so we replace them with

$$\Lambda(f_1, \ldots, f_n) := \frac{1}{2} \int \frac{(f_1(x_1) - f_1(x_1^*)) f_2(x_2) \cdots f_n(x_n)}{\det \left(\begin{smallmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{smallmatrix} \right)} \, \mathrm{d}x_1 \, \mathrm{d}x_2 \ldots \mathrm{d}x_n \tag{6.3}$$

where x_1^* is the reflection of x_1 in the hyperplane determined by the other variables and

$$\Lambda_S(f_1, \ldots, f_n) := \frac{1}{2} \int \frac{(f_1(\omega_1) - f_1(\omega_1^*)) f_2(\omega_2) \cdots f_n(\omega_n)}{\det(\omega_1, \ldots, \omega_n)} \, \mathrm{d}\omega_1 \, \mathrm{d}\omega_2 \ldots \mathrm{d}\omega_n \tag{6.4}$$

where ω_1^* is the reflection of ω_1 in the great hypercircle determined by by the other variables.

As a purely formal exercise we can calculate

$$\begin{aligned}\Lambda(f_1,\dots,f_n) =&\frac{1}{2}\int \frac{(f_1(x_1)-f_1(x_1^*))f_2(x_2)\cdots f_n(x_n)}{\det\left(\begin{smallmatrix}1 & 1 & \dots & 1\\ x_1 & x_2 & \dots & x_n\end{smallmatrix}\right)}\,\mathrm{d}x_1\,\mathrm{d}x_2\dots\mathrm{d}x_n\\ =&\frac{1}{2}\int \frac{f_1(x_1)f_2(x_2)\cdots f_n(x_n)}{\det\left(\begin{smallmatrix}1 & 1 & \dots & 1\\ x_1 & x_2 & \dots & x_n\end{smallmatrix}\right)}\,\mathrm{d}x_1\,\mathrm{d}x_2\dots\mathrm{d}x_n\\ &-\frac{1}{2}\int \frac{f_1(x_1^*)f_2(x_2)\cdots f_n(x_n)}{\det\left(\begin{smallmatrix}1 & 1 & \dots & 1\\ x_1 & x_2 & \dots & x_n\end{smallmatrix}\right)}\,\mathrm{d}x_1\,\mathrm{d}x_2\dots\mathrm{d}x_n\\ =&\int \frac{f_1(x_1)f_2(x_2)\cdots f_n(x_n)}{\det\left(\begin{smallmatrix}1 & 1 & \dots & 1\\ x_1 & x_2 & \dots & x_n\end{smallmatrix}\right)}\,\mathrm{d}x_1\,\mathrm{d}x_2\dots\mathrm{d}x_n\end{aligned}$$

since the change of variables $x_1 \mapsto x_1^*$ has Jacobian 1, $x_1^{**} = x_1$ and

$$\det\begin{pmatrix}1 & 1 & \dots & 1\\ x_1 & x_2 & \dots & x_n\end{pmatrix} = -\det\begin{pmatrix}1 & 1 & \dots & 1\\ x_1^* & x_2 & \dots & x_n\end{pmatrix}$$

which follows by noting that the determinants are the signed volumes of the simplices whose vertices are $x_1,\dots,x_n$ and $x_1^*, x_2,\dots,x_n$ respectively and these simplices have the same unsigned volume but different orientations.

The following lemma establishes that (6.3) and (6.4) are sensible definitions.

Lemma 6.1. *Let $f_1,\dots,f_n$ be functions in $C_c^\infty(\mathbb{R}^{n-1})$ or $C^\infty(S^{n-1})$ respectively. Then*

1. *the integrals in* (6.3) *or* (6.4) *are absolutely convergent,*
2. *the numerator $(f_1(x_1)-f_1(x_1^*))f_2(x_2)\cdots f_n(x_n)$ in* (6.3) *can be replaced by*

   $$f_1(x_1)\cdots f_{i-1}(x_{i-1})(f_i(x_i)-f_i(x_i^*))f_{i+1}(x_{i+1})\cdots f_n(x_n)$$

   for any $i = 1,\dots,n$ without affecting the value of the integral and
3. *the numerator $(f_1(\omega_1)-f_1(\omega_1^*))f_2(\omega_2)\cdots f_n(\omega_n)$ in* (6.4) *can be replaced by*

   $$f_1(\omega_1)\cdots f_{i-1}(\omega_{i-1})(f_i(\omega_i)-f_i(\omega_i^*))f_{i+1}(\omega_{i+1})\cdots f_n(\omega_n)$$

   for any $i = 1,\dots,n$ without affecting the value of the integral.

The symbols x_i^* and ω_i^* have the obvious meaning analogous to x_1^* and ω_1^*. We shall postpone the proof of this lemma to the end of this section.

What we are interested in are estimates of the form

$$|\Lambda(f_1,\dots,f_n)| \lesssim \|f_1\|_{L^{p_1}(\mathbb{R}^{n-1})}\cdots\|f_n\|_{L^{p_n}(\mathbb{R}^{n-1})} \tag{6.5}$$

and

$$|\Lambda_S(f_1,\ldots,f_n)| \lesssim \|f_1\|_{L^{p_1}(S^{n-1})}\cdots\|f_n\|_{L^{p_n}(S^{n-1})} \tag{6.6}$$

where $A \lesssim B$ denotes that there is an absolute constant C, depending only on the dimension, such that $A \le CB$. We shall prove the following theorems.

Theorem 6.2. *Let $\mathcal{S}$ be the closed polytope in $\mathbb{R}^n$ whose vertices are the n permutations of the n-tuple $(\frac{n-2}{n-1},\ldots,\frac{n-2}{n-1},1)$. Then (6.5) holds if and only if $(\frac{1}{p_1},\ldots,\frac{1}{p_n})$ lies in the interior of $\mathcal{S}$, relative to the hyperplane that $\mathcal{S}$ lies in. For $n \ge 3$, the estimate holds on the boundary of $\mathcal{S}$ if each f_j with j for which $\frac{1}{p_j} = \frac{n-2}{n-1}$ is restricted to be a characteristic function of a set but the other f_j's may be unrestricted. The estimate fails if any f_j with j for which $\frac{1}{p_j} = \frac{n-2}{n-1}$ is taken unrestricted.*

Remark 6.3. Each point (q_j) in $\mathcal{S}$ lies in the hyperplane Π defined by the equation

$$\sum_{j=1}^{n} q_j = n-1.$$

When we speak about the exterior, the interior and the boundary of $\mathcal{S}$ we understand it to be taken relative to Π.

For (6.6) we have the following.

Theorem 6.4. *Let $\tilde{\mathcal{S}}$ be the half-open polytope in the closed first 2^n-tant in $\mathbb{R}^n$ consisting of points $\bar{p}$ which have the form $\bar{p} = \bar{q} - a$ with $\bar{q}$ in the interior of $\mathcal{S}$ and a a vector in $\mathbb{R}^n$, all of whose entries are non-negative. Then (6.6) holds if $(\frac{1}{p_1},\ldots,\frac{1}{p_n})$ lies in $\tilde{\mathcal{S}}$. For $n \ge 3$, the estimate holds when each f_j for j in a set $J \subset \{1,\ldots,n\}$ is restricted to be a characteristic function of a set if $\bar{p} = \left(\frac{1}{p_1},\ldots,\frac{1}{p_n}\right)$ has the form $\bar{p} = \bar{q} - a$ with a as before and $\bar{q} = (q_j)$ on the boundary of $\mathcal{S}$ and $q_j \neq \frac{n-2}{n-1}$ for $j \notin J$.*

Remark 6.5. We firmly believe that the condition of this theorem is also a necessary condition for (6.6) to hold but we have not been able to verify that. See however also Remark 6.13 below.

By specialising these theorems to the centre of $\mathcal{S}$ we get:

Corollary 6.6.

$$|\Lambda(f_1,\ldots,f_n)| \lesssim \prod_{i=1}^{n} \|f_i\|_{L^{\frac{n}{n-1}}(\mathbb{R}^{n-1})}. \tag{6.7}$$

Corollary 6.7.

$$|\Lambda_S(f_1,\ldots,f_n)| \lesssim \prod_{i=1}^{n} \|f_i\|_{L^{\frac{n}{n-1}}(S^{n-1})}. \tag{6.8}$$

For $n \geq 3$ the positive results of Theorem 6.2 follow from the following estimate.

Theorem 6.8. *Let $n \geq 3$ and $\chi_{E_1}, \ldots, \chi_{E_{n-1}}$ be characteristic functions of $n-1$ measurable sets in $\mathbb{R}^{n-1}$ and f_n be a measurable function on $\mathbb{R}^{n-1}$. Then*

$$|\Lambda(\chi_{E_1}, \ldots, \chi_{E_{n-1}}, f_n)| \lesssim \|\chi_{E_1}\|_{\frac{n-1}{n-2}} \cdots \|\chi_{E_{n-1}}\|_{\frac{n-1}{n-2}} \|f_n\|_1. \tag{6.9}$$

Let us note how we can use multilinear interpolation to pass from this estimate to the general result of the theorem. Firstly note that convexity gives directly that (6.5) holds for tuples $(\frac{1}{p_1}, \ldots, \frac{1}{p_n})$ in $\mathcal{S}$ if each f_j is restricted to be a characteristic function f_j. Now take an element $\bar{p} \in \mathcal{S}$ and assume that $\frac{1}{p_j} = \frac{n-2}{n-1}$ if and only if $j \leq k$ where $k \leq n$. Let us fix sets E_j for $1 \leq j \leq k$ and note that we have

$$|\Lambda(\chi_{E_1}, \ldots, \chi_{E_n})| \lesssim \prod_{j \leq k} \|\chi_{E_j}\|_{\frac{n-1}{n-2}} \prod_{j > k} \|\chi_{E_j}\|_{q_j} \tag{6.10}$$

if each q_j is sufficiently close to p_j and

$$\sum_{j=k+1}^{n} \frac{1}{q_j} = 1 + \frac{n-2}{n-1}(n-k-1).$$

This shows that we can use Marcinkiewicz interpolation, see for example [21], to strengthen this result to

$$|\Lambda(\chi_{E_1}, \ldots, \chi_{E_k}, f_{k+1}, \ldots, f_n)| \lesssim \prod_{j \leq k} \|\chi_{E_j}\|_{\frac{n-1}{n-2}} \prod_{j > k} \|f_j\|_{p_j}. \tag{6.11}$$

By permuting the indices we arrive at the estimate of the theorem.

The remaining parts of Theorem 6.2 can be seen from examples which we now present.

Example 6.9. Let us assume that inequality (6.5) holds for the dilated functions $\phi_1(\frac{\cdot}{R}), \ldots, \phi_n(\frac{\cdot}{R})$ for all $R > 0$. Then

$$\left| \int \frac{\phi_1(\frac{x_1}{R}), \ldots, \phi_n(\frac{x_n}{R})}{\det\left(\begin{smallmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{smallmatrix}\right)} \mathrm{d}x_1 \, \mathrm{d}x_2 \ldots \mathrm{d}x_n \right| \lesssim \|\phi_1(\frac{\cdot}{R})\|_{p_1} \cdots \|\phi_n(\frac{\cdot}{R})\|_{p_n}$$

so

$$\left| \int \frac{\phi_1(\frac{x_1}{R}), \ldots, \phi_n(\frac{x_n}{R})}{\det\left(\begin{smallmatrix} 1 & 1 & \cdots & 1 \\ \frac{x_1}{R} & \frac{x_2}{R} & \cdots & \frac{x_n}{R} \end{smallmatrix}\right) \cdot R^{n-1}} \frac{\mathrm{d}x_1}{R^{n-1}} \frac{\mathrm{d}x_2}{R^{n-1}} \cdots \frac{\mathrm{d}x_n}{R^{n-1}} \cdot R^{n(n-1)} \right|$$

$$\lesssim \left(\prod_{i=1}^{n} R^{\frac{n-1}{p_i}} \right) \|\phi_1\|_{p_1} \cdots \|\phi_n\|_{p_n}$$

so

$$R^{(n-1)^2} \lesssim R^{\sum (n-1)\frac{1}{p_i}}$$

so

$$\sum_{i=1}^{n} \frac{1}{p_i} = n - 1. \tag{6.12}$$

Example 6.10. As stated above we get for $n = 2$ that $\Lambda(f, g) = \pi\langle Hf, g\rangle$ where Hf is the Hilbert transform of f. Thus by well-known properties we see that

$$\begin{gathered} |\Lambda(f,g)| \lesssim \|f\|_{p_1} \|g\|_{p_2} \\ \text{if} \quad \frac{1}{p_1} + \frac{1}{p_2} = 1 \quad \text{and} \quad 1 < p_1, p_2 < \infty. \end{gathered} \tag{6.13}$$

Aside from the endpoints, this is the best estimate we could hope for in the light of the previous example.

Example 6.11. When $n \geq 3$ there is a further restriction on the values of p_j for which (6.5) can hold.

To see this let us first of all note that there exist non-empty open cones $\mathcal{C}_1, \ldots, \mathcal{C}_n$ with vertices at the origin in $\mathbb{R}^{n-1}$ such that if $x_1 \in \mathcal{C}_1, \ldots, x_n \in \mathcal{C}_n$ then

$$\det \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} > 0.$$

To construct these cones we can for example take $\mu_1, \ldots, \mu_n \in S^{n-1}$ to be the vertices of a regular simplex with centre at the origin. We shall denote the simplex whose vertices are $\nu_1, \ldots, \nu_n$ by $\mathcal{T}_{(\nu_i)}$. Now, the signed volume of $\mathcal{T}_{(\mu_i)}$ is given by

$$\det \begin{pmatrix} 1 & \cdots & 1 \\ \mu_1 & \cdots & \mu_n \end{pmatrix} \tag{6.14}$$

which can therefore not equal zero and we may furthermore assume that we have carried out the numbering of the μ's in such a way that this determinant is positive.

Let us note that if the origin lies in the interior of a simplex $\mathcal{T}_{(\nu_i)}$ then it also lies in the interior of $\mathcal{T}_{(r_i\nu_i)}$ for any positive scalars r_i. We can prove this iteratively if we know that this holds when all of the r_i's except one equal 1. We may then further assume that this exceptional r_i is r_1.

Now, the origin lies in the interior of $\mathcal{T}_{(\nu_i)}$ if and only if the line connecting ν_1 and the origin intersects the interior of the facet opposite ν_1 and this intersection lies beyond the origin. If we replace ν_1 by $r\nu_1$ for $r > 0$ then this line and the opposite facet remain unaltered and the intersection will still lie beyond the origin.

Now let M_i be a small neighbourhood in S^{n-1} around ν_i such that for any tuple $(\tilde{\mu}_i)$ in $M_1 \times \cdots \times M_n$ we have that the determinant in (6.14) is positive and that the origin lies in the interior of the simplex $\mathcal{T}_{(\tilde{\mu}_i)}$.

By what we have said it is now clear that we may take $\mathcal{C}_i$ to be the smallest cone with vertex at the origin which contains M_i.

With this set-up in hand we let $\phi_1, \dots, \phi_n$ be non-negative C_c^∞ functions such that $\operatorname{supp} \phi_i \subset \mathcal{C}_i$. We also insist that ϕ_i is supported in $|x| < \frac{1}{10}$ for $i = 1, \dots, k$ while ϕ_i is supported in $\frac{1}{2} < |x| < 1$ for $i = k+1, \dots, n$. Here k is an integer between 1 and n. These conditions will continue to hold if we replace all the ϕ_i's for $i \leq k$ by $\phi_{i,\epsilon} : x \mapsto \phi_i(\frac{x}{\epsilon})$ for $\epsilon < 1$. Now,

$$\det \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} = \det(x_2 - x_1, \cdots, x_n - x_1) \leq |x_2 - x_1| \cdots |x_n - x_1|$$

by Hadamard's theorem so

$$\begin{aligned} &\Lambda(\phi_{1,\epsilon}, \dots, \phi_{k,\epsilon}, \phi_{k+1}, \dots, \phi_n) \\ &\qquad \gtrsim \int \cdots \int \frac{\phi_1(\frac{x_1}{\epsilon}) \dots \phi_k(\frac{x_k}{\epsilon}) \phi_{k+1}(x_{k+1}) \phi_n(x_n)}{|x_2 - x_1| \dots |x_k - x_1| |x_{k+1}| \dots |x_n|} \, \mathrm{d}x_1 \dots \mathrm{d}x_n \end{aligned}$$

because we have $|x_i - x_1| \sim |x_i|$ for all $i > k$. We then have

$$\begin{aligned} &\Lambda(\phi_{1,\epsilon}, \dots, \phi_{k,\epsilon}, \phi_{k+1}, \dots, \phi_n) \\ &\gtrsim \epsilon^{(n-1)k-(k-1)} \int \cdots \int \frac{\phi_1(\frac{x_1}{\epsilon}) \dots \phi_k(\frac{x_k}{\epsilon}) \phi_{k+1}(x_{k+1}) \dots \phi_n(x_n)}{|\frac{x_2}{\epsilon} - \frac{x_1}{\epsilon}| \dots |\frac{x_k}{\epsilon} - \frac{x_1}{\epsilon}| |x_{k+1}| \dots |x_n|} \\ &\qquad\qquad\qquad\qquad \frac{\mathrm{d}x_1}{\epsilon^{n-1}} \cdots \frac{\mathrm{d}x_k}{\epsilon^{n-1}} \, \mathrm{d}x_{k+1} \dots \mathrm{d}x_n \\ &\gtrsim \epsilon^{(n-1)k-(k-1)} \int \cdots \int \frac{\phi_1(x_1) \dots \phi_n(x_n)}{|x_2 - x_1| \dots |x_k - x_1| |x_{k+1}| \dots |x_n|} \, \mathrm{d}x_1 \dots \mathrm{d}x_n \end{aligned}$$

and

$$\prod_{i=1}^{k} \|\phi_{i,\epsilon}\|_{p_i} = \epsilon^{(n-1)\sum_{i=1}^{k} \frac{1}{p_i}} \prod_{i=1}^{n-1} \|\phi_i\|_{p_i}$$

so we must have

$$\epsilon^{(n-1)k-(k-1)} \lesssim \epsilon^{(n-1)\sum_{i=1}^{k} \frac{1}{p_i}} \quad \text{for } \epsilon < 1$$

so

$$\frac{(n-1)k - (k-1)}{n-1} \geq \sum_{i=1}^{k} \frac{1}{p_i}.$$

In particular, for $k = n-1$, this tells us that

$$\frac{n^2 - 3n + 3}{n-1} \geq \sum_{i=1}^{n-1} \frac{1}{p_i}$$

and this together with (6.12) and renaming of the variables gives us that

$$\frac{1}{p_i} \geq \frac{n-2}{n-1} \tag{6.15}$$

for all $i = 1, \dots, n$. The polyhedron defined by (6.12) and (6.15) has the permutations of the n-tuple $\left(\frac{n-2}{n-1}, \dots, \frac{n-2}{n-1}, 1\right)$ as vertices so we see that (6.5) can only hold at points in $\mathcal{S}$.

Example 6.12. Let us see that we cannot hope to strengthen the estimates on the boundary of $\mathcal{S}$ to strong-type estimates.

We let $\mathcal{C}_i$ be as in the previous example and take ϕ_i to be non-negative functions supported in $\mathcal{C}_i$. Assume that ϕ_i is supported in $|x| < 1$ for $i < n$ and ϕ_n is supported in $|x| > 10$. As before we can estimate by Hadamard's theorem and get

$$\Lambda(\phi_1, \dots, \phi_n) \gtrsim \int \cdots \int \frac{\phi_1(x_1) \cdots \phi_n(x_n)}{|x_2 - x_1| \cdots |x_{n-1} - x_1||x_n|} \, \mathrm{d}x_1 \cdots \mathrm{d}x_n. \tag{6.16}$$

Let us now assume that ϕ_n has the form $\phi_n(x) = \phi_\omega(\omega)\phi_r(r)$ with $x = r\omega$ in polar coordinates where $\phi_r(r) = (r^{n-2} \log r)^{-1}$ for $10 < r < b$. Then the right hand side of (6.16) contains a factor larger than

$$\int_{10}^{b} \frac{1}{(r^{n-2} \log r) r} r^{n-2} \, \mathrm{d}r = \int_{10}^{b} \frac{1}{\log r} \frac{\mathrm{d}r}{r} = \log b - \log 10.$$

On the other hand we see that

$$\|\phi_n\|_{\frac{n-1}{n-2}} = C \left(\int_{10}^{b} \left(\frac{1}{r^{n-2} \log r} \right)^{\frac{n-1}{n-2}} r^{n-2} \, \mathrm{d}r \right)^{\frac{n-2}{n-1}} = C \left(\int_{10}^{b} \left(\frac{1}{\log r} \right)^{\frac{n-1}{n-2}} \frac{\mathrm{d}r}{r} \right)^{\frac{n-2}{n-1}}$$

which is less than a constant independent of b. Since b can be arbitrarily large we get a contradiction unless

$$\frac{1}{p_j} > \frac{n-2}{n-1}.$$

Remark 6.13. It is clear that we can adapt Example 6.9 for $R < 1$ and Example 6.11 to the form Λ_S. The former of these shows that (6.6) can only hold if $\bar{p} = (\frac{1}{p_1}, \dots, \frac{1}{p_n})$ lies in the intersection of the first 2^n-tant with the closed half-space whose boundary is Π as defined in Remark 6.3 and which contains the origin.

The latter example shows that if $\bar{p}$ lies in Π then (6.6) cannot hold if $\bar{p}$ lies in the exterior of $\mathcal{S}$. On the other hand Theorem 6.4 shows that (6.6) holds if $\bar{p}$ is in the interior of $\mathcal{S}$.

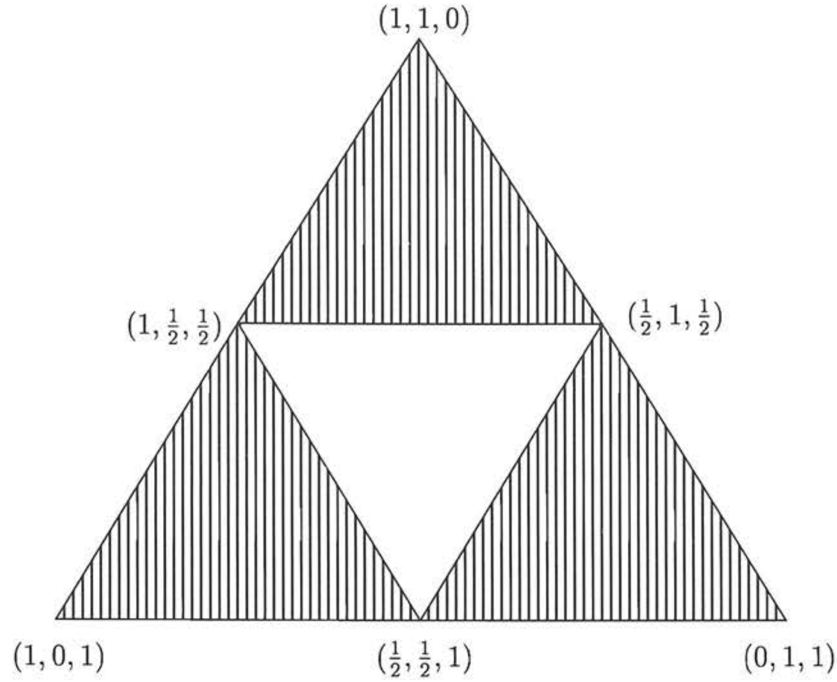


Figure 6.1: $n = 3$ The estimate (6.5) holds in the open unshaded region and fails in the open shaded region.

Proof of Theorem 6.8. The proof will be based on Theorem 6.4 and Lemma 6.14 below. First of all we note that

$$\begin{aligned}
|\Lambda(\chi_{E_1}, \dots, \chi_{E_{n-1}}, f_n)| &= \left| \int \frac{\chi_{E_1}(x_1) \cdots \chi_{E_{n-1}}(x_{n-1}) f_n(x_n)}{\det \left(\begin{smallmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{smallmatrix} \right)} \, \mathrm{d}x_1 \cdots \mathrm{d}x_n \right| \\
&\le \|f_n\|_1 \sup_{x_n} \left| \int \frac{\chi_{E_1}(x_1) \cdots \chi_{E_{n-1}}(x_{n-1})}{\det \left(\begin{smallmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{smallmatrix} \right)} \, \mathrm{d}x_1 \cdots \mathrm{d}x_{n-1} \right| \\
&\le \|f_n\|_1 \sup_{x_n} \left| \int \frac{\chi_{E_1}(x_1) \cdots \chi_{E_{n-1}}(x_{n-1})}{\det \left(\begin{smallmatrix} 1 & \cdots & 1 & 1 \\ x_1 - x_n & \cdots & x_{n-1} - x_n & 0 \end{smallmatrix} \right)} \, \mathrm{d}x_1 \cdots \mathrm{d}x_{n-1} \right| \\
&= \|f_n\|_1 \sup_{x_n} \left| \int \frac{\chi_{\tilde{E}_1}(x_1) \cdots \chi_{\tilde{E}_{n-1}}(x_{n-1})}{\det(x_1, \dots, x_{n-1})} \, \mathrm{d}x_1 \cdots \mathrm{d}x_{n-1} \right|
\end{aligned}$$

where $\tilde{E}_i := E_i - x_n$. Since $\|\chi_{\tilde{E}_i}\|_p = \|\chi_{E_i}\|_p$ we will drop these tildes. Let us then define $\tilde{\Lambda}(\chi_{E_1}, \dots \chi_{E_{n-1}})$ to be the quantity inside the modulus signs on the far right hand side of the last chain of inequalities. We change to polar coordinates, $x_i = r_i \omega_i$ with $r_i \in \mathbb{R}_+$ and $\omega_i \in S^{n-2}$. Then $\det(x_1, \dots, x_{n-1}) = r_1 \cdots r_{n-1} \det(\omega_1, \dots, \omega_{n-1})$ and $\mathrm{d}x_i = r_i^{n-2} \, \mathrm{d}r_i \, \mathrm{d}\omega_i$ ($\mathrm{d}\omega_i$ is the unnormalised induced Lebesgue measure on the sphere) so

$$
\begin{aligned}
\tilde{\Lambda}(\chi_{E_1},\dots,\chi_{E_{n-1}}) &= \int \frac{\chi_{E_1}(r_1\omega_1)\cdots\chi_{E_{n-1}}(r_{n-1}\omega_{n-1})}{r_1\cdots r_{n-1}\det(\omega_1,\dots,\omega_{n-1})} \\
&\qquad \cdot (r_1\cdots r_{n-1})^{n-2}\,\mathrm{d}r_1\cdots\mathrm{d}r_{n-1}\,\mathrm{d}\omega_1\cdots\mathrm{d}\omega_{n-1} \\
&= \int \frac{F_{n-1}(\chi_{E_1})(\omega_1)\cdots F_{n-1}(\chi_{E_{n-1}})(\omega_{n-1})}{\det(\omega_1,\dots,\omega_{n-1})}\,\mathrm{d}\omega_1\cdots\mathrm{d}\omega_{n-1} \\
&= \Lambda_S(F_{n-1}(\chi_{E_1}),\dots,F_{n-1}(\chi_{E_{n-1}})) \qquad (6.17)
\end{aligned}
$$

where $F_{n-1}(f)(\omega) = \int_{\mathbb{R}_+} f(r\omega)r^{n-3}\,\mathrm{d}r$ and in (6.17) we have that Λ_S acts on functions on S^{n-2}. Thus we have separated $\tilde{\Lambda}$ into a radial part, F_{n-1}, and an angular part. By Theorem 6.4 we can estimate (6.17) by a constant multiple of

$$\|F_{n-1}(\chi_{E_1})\|_{L^{\frac{n-1}{n-2}}(S^{n-2})}\cdots\|F_{n-1}(\chi_{E_{n-1}})\|_{L^{\frac{n-1}{n-2}}(S^{n-2})}$$

so Theorem 6.2 will follow from the following lemma. □

Lemma 6.14.

$$\|F_{n-1}(\chi_E)\|_{L^{\frac{n-1}{n-2}}(S^{n-2})} \lesssim \|\chi_E\|_{L^{\frac{n-1}{n-2}}(\mathbb{R}^{n-1})}. \qquad (6.18)$$

Remark 6.15. We note that the estimate in this lemma does not hold for general functions as can be seen by testing on the function $f(r\omega) = (r^{n-2}\log r)^{-1}$ similarly to Example 6.12.

Proof of Lemma 6.14. If $n = 3$ we want to prove that

$$\left\|\int_{\mathbb{R}_+} \chi_E(r\omega)\,\mathrm{d}r\right\|_{L^2(S^1)} \lesssim \|\chi_E\|_{L^2(\mathbb{R}^2)}$$

which is equivalent to

$$\int_{S^1}\left|\int_{\mathbb{R}_+}\chi_E(r\omega)\,\mathrm{d}r\right|^2\mathrm{d}\omega \lesssim \int_{S^1}\int_{\mathbb{R}_+}|\chi_E(r\omega)|^2 r\,\mathrm{d}r\,\mathrm{d}\omega.$$

Define $E_\omega := \{r \in \mathbb{R}_+ : r\omega \in E\}$. We see that it is enough to prove that $|E_\omega|^2 \lesssim \int_{E_\omega} r\,\mathrm{d}r$ holds for each $\omega \in S^1$. The left hand side in this inequality depends only on the measure of E_ω and the infimum of the right hand side, for sets of fixed measure, is clearly attained when $E_\omega = [0, |E_\omega|]$. In this case $\int_{E_\omega} r\,\mathrm{d}r = \frac{1}{2}|E_\omega|^2$ so $|E_\omega|^2 \le 2\int_{E_\omega} r\,\mathrm{d}r$.

More generally, the same reasoning shows that $|E_\omega|^m \lesssim \int_{E_\omega} r^{m-1}\,\mathrm{d}r$. It follows that

$$\begin{aligned}
\int_{E_\omega} r^{n-3}\,\mathrm{d}r &\le \left(\int_{E_\omega} (r^{n-3})^{\frac{n-2}{n-3}}\,\mathrm{d}r\right)^{\frac{n-3}{n-2}} \left(\int_{E_\omega} \mathrm{d}r\right)^{1-\frac{n-3}{n-2}} \qquad \text{(by Hölder)}\\
&= \left(\int_{E_\omega} r^{n-2}\,\mathrm{d}r\right)^{\frac{n-3}{n-2}} \left(\int_{E_\omega} \mathrm{d}r\right)^{\frac{1}{n-2}}\\
&\lesssim \left(\int_{E_\omega} r^{n-2}\,\mathrm{d}r\right)^{\frac{n-3}{n-2}} \left(\int_{E_\omega} r^{n-2}\,\mathrm{d}r\right)^{\frac{1}{(n-2)(n-1)}}\\
&= \left(\int_{E_\omega} r^{n-2}\,\mathrm{d}r\right)^{\frac{n-2}{n-1}}
\end{aligned}$$

which is to say that

$$\left(\int_{E_\omega} r^{n-3}\,\mathrm{d}r\right)^{\frac{n-1}{n-2}} \lesssim \int_{E_\omega} r^{n-2}\,\mathrm{d}r.$$

Then we see that

$$\int_{S^{n-2}} \left|\int_{\mathbb{R}_+} \chi_E(r\omega) r^{n-3}\,\mathrm{d}r\right|^{\frac{n-1}{n-2}} \mathrm{d}\omega \lesssim \int_{S^{n-2}}\int_{\mathbb{R}_+} \chi_E(r\omega) r^{n-2}\,\mathrm{d}r\,\mathrm{d}\omega$$

so

$$\|F_{n-1}(\chi_E)\|_{L^{\frac{n-1}{n-2}}(S^{n-2})} \lesssim \|\chi_E\|_{L^{\frac{n-1}{n-2}}(\mathbb{R}^{n-1})}.$$

This completes the proof of the lemma. □

Proof of Theorem 6.4. For $n = 2$ we see that

$$\Lambda_S(f_1, f_2) = \int_{S^1}\int_{S^1} \frac{f_1(\omega_1) f_2(\omega_2)}{\sin(\omega_1 - \omega_2)} \lesssim \|f_1\|_{L^{p_1}(S^1)} \|f_2\|_{L^{p_2}(S^1)}$$

provided that $p_1, p_2 > 1$ and $\frac{1}{p_1} + \frac{1}{p_2} = 1$ since $(\sin(\omega_1 - \omega_2))^{-1} = \frac{1}{2}\tan\frac{1}{2}(\omega_1 - \omega_2) + \frac{1}{2}\cot\frac{1}{2}(\omega_1 - \omega_2)$ so the left hand side is the sum of two Hilbert transforms and the result is known.

So that we have a clearer relation with the proof of Theorem 6.8 we shall now change our indexing and in effect increase n by one. We will proceed by using induction and will assume that we have some $n \ge 4$ and that we have proved Corollary 6.7 on S^{n-3}, that is

$$|\Lambda_S(f_1, \ldots, f_{n-2})| \lesssim \|f_1\|_{L^{\frac{n-2}{n-3}}(S^{n-3})} \cdots \|f_{n-2}\|_{L^{\frac{n-2}{n-3}}(S^{n-3})} \tag{6.19}$$

and we are interested in proving

$$|\Lambda_S(f_1, \ldots, f_{n-1})| \lesssim \|f_1\|_{L^{p_1}(S^{n-2})} \cdots \|f_{n-1}\|_{L^{p_{n-1}}(S^{n-2})} \tag{6.20}$$

with $\left(\frac{1}{p_1}, \ldots, \frac{1}{p_{n-1}}\right)$ in the interior of $\mathcal{S}$ (with n replaced by $n-1$). Again by multilinear interpolation, it is enough to prove the estimate for characteristic functions at the vertex ($n-1$-tuple) $\left(\frac{n-3}{n-2},\ldots,\frac{n-3}{n-2},1\right)$. We proceed in the following manner. By definition, $\Lambda_S(f_1,\ldots,f_{n-1})$ equals

$$\frac{1}{2}\int \frac{(f_1(\omega_1)-f_1(\omega_1^*))f_2(\omega_2)\cdots f_{n-1}(\omega_{n-1})}{\det(\omega_1,\ldots,\omega_{n-1})}\,\mathrm{d}\omega_1\cdots\mathrm{d}\omega_{n-1} \tag{6.21}$$

where ω_1^* is the reflection of ω_1 in the great hypercircle containing ω_2 up to ω_{n-1}. We bound this by

$$\|f_{n-1}\|_{L^1(S^{n-2})}\sup_{\omega_{n-1}}\left|\frac{1}{2}\int \frac{(f_1(\omega_1)-f_1(\omega_1^*))f_2(\omega_2)\cdots f_{n-2}(\omega_{n-2})}{\det(\omega_1,\ldots,\omega_{n-1})}\,\mathrm{d}\omega_1\cdots\mathrm{d}\omega_{n-2}\right|$$

We thus want to show that

$$\sup_{\omega_{n-1}}\left|\frac{1}{2}\int \frac{(f_1(\omega_1)-f_1(\omega_1^*))f_2(\omega_2)\cdots f_{n-2}(\omega_{n-2})}{\det(\omega_1,\ldots,\omega_{n-1})}\,\mathrm{d}\omega_1\cdots\mathrm{d}\omega_{n-2}\right| \lesssim \|f_1\|_{L^{\frac{n-2}{n-3}}(S^{n-2})}\cdots\|f_{n-2}\|_{L^{\frac{n-2}{n-3}}(S^{n-2})}$$

holds for all f_j being characteristic functions.

By rotational invariance, we can take ω_{n-1} to be the north pole $\binom{1}{0}$. We then split the integral in each of the variables $\omega_1\ldots\omega_{n-2}$ into two integrals, one over each hemisphere.

Because

$$\det(\omega_1,\ldots,-\omega_i,\ldots,\omega_{n-1}) = -\det(\omega_1,\ldots,\omega_i,\ldots,\omega_{n-1})$$

it is enough to consider the integral over the northern hemispheres

$$S_+^{n-2} := \{\omega_0=(\omega_{01},\ldots,\omega_{0,n-1})\in S^{n-2}:\omega_{01}>0\}.$$

Since ω_1^* is the reflection of ω_1 in a great hypercircle containing the north pole we see that ω_1 and ω_1^* will always lie in the same hemisphere. To work with the integral over S_+^{n-2} we change variables from S_+^{n-2} to $\{1\}\times\mathbb{R}^{n-2}$. Specifically, we write $\omega_0\in S_+^{n-2}$ as $(\cos\theta_0,\widetilde{\omega}_0\sin\theta_0)$ where $0\le\theta_0<\frac{\pi}{2}$ and $\widetilde{\omega}_0\in S^{n-3}$. Define

$$\psi(\omega_0) := \frac{1}{\cos\theta_0}(\cos\theta_0,\widetilde{\omega}_0\sin\theta_0) = (1,\widetilde{\omega}_0\tan\theta_0).$$

Since $\widetilde{\omega}_0\in S^{n-3}$, the expression $\widetilde{\omega}_0\sin\theta_0$ for a fixed $\widetilde{\omega}_0$ parametrises an $(n-3)$ dimensional sphere of radius $\sin\theta_0$ and the expression $\widetilde{\omega}_0\tan\theta_0$ parametrises a similar sphere of radius $\tan\theta_0$. This contributes a factor

$$\left(\frac{\sin\theta_0}{\tan\theta_0}\right)^{n-3} = \cos^{n-3}\theta_0$$

to $(J\psi^{-1})(\psi(\omega_0))$. Also,

$$\frac{\partial\psi}{\partial\theta_0}(\omega_0) = \frac{\partial}{\partial\theta_0}\tan\theta_0 = \frac{1}{\cos^2\theta_0}$$

so $(J\psi^{-1})(\psi(\omega_0)) = \cos^{n-1}\theta_0$. The integral (6.21) thus becomes

$$\int_{(\{1\}\times\mathbb{R}^{n-2})^{n-2}} \frac{f_1(\psi^{-1}(\widetilde{x}_1))\dots f_{n-2}(\psi^{-1}(\widetilde{x}_{n-2}))}{\det\left(\frac{\widetilde{x}_1}{|\widetilde{x}_1|},\dots,\frac{\widetilde{x}_{n-2}}{|\widetilde{x}_{n-2}|}\begin{smallmatrix}1\\0\end{smallmatrix}\right)} \left(\prod_{i=1}^{n-2}\cos\theta_{i0}\right)^{n-1} \mathrm{d}\widetilde{x}_1\cdots\mathrm{d}\widetilde{x}_{n-2}$$

and we can pull $|\widetilde{x}_i| = \sqrt{1+\tan^2\theta_{i0}} = (\cos\theta_{i0})^{-1}$ out of the determinant.

Let $\widetilde{x}_i = \binom{1}{y_i}$ and $\psi^{-1}(\widetilde{x}_i) = \widetilde{\psi}^{-1}(y_i)$. Then since

$$\cos\theta_{i0} = \frac{1}{|\widetilde{x}_i|} = \frac{1}{(1+|y_i|^2)^{\frac{1}{2}}}$$

we see that the integral becomes

$$\begin{aligned}
&\int_{(\mathbb{R}^{n-2})^{n-2}} \frac{f_1(\widetilde{\psi}^{-1}(y_1))\dots f_{n-2}(\widetilde{\psi}^{-1}(y_{n-2}))}{\det(y_1,\dots,y_{n-2})} \prod_{i=1}^{n-2}\frac{1}{(1+|y_i|^2)^{\frac{n-2}{2}}}\,\mathrm{d}y_1\cdots\mathrm{d}y_{n-2}\\
&= \int_{(S^{n-3})^{n-2}}\int_{(\mathbb{R}_+)^{n-2}} \frac{f_1(\widetilde{\psi}^{-1}(\widetilde{r}_1\widetilde{\omega}_1))\dots f_{n-2}(\widetilde{\psi}^{-1}(\widetilde{r}_{n-2}\widetilde{\omega}_{n-2}))}{\widetilde{r}_1\dots\widetilde{r}_{n-2}\det(\widetilde{\omega}_1,\dots,\widetilde{\omega}_{n-2})}\\
&\qquad \prod_{i=1}^{n-2}\frac{1}{(1+\widetilde{r}_i^2)^{\frac{n-2}{2}}}(\widetilde{r}_1\dots\widetilde{r}_{n-2})^{n-3}\,\mathrm{d}\widetilde{r}_1\cdots\mathrm{d}\widetilde{r}_{n-2}\,\mathrm{d}\widetilde{\omega}_1\cdots\mathrm{d}\widetilde{\omega}_{n-2}
\end{aligned}$$

where we have changed to polar coordinates again. By the induction hypothesis we can estimate the angular part of this by

$$\prod_{i=1}^{n-2}\left\|\int\frac{f_i(\widetilde{\psi}^{-1}(r\cdot))}{(1+r)^{\frac{n-2}{2}}}r^{n-4}\,\mathrm{d}r\right\|_{L^{\frac{n-2}{n-3}}(S^{n-3})}.$$

We want to bound this by

$$\prod_{i=1}^{n-2}\|f_i\|_{L^{\frac{n-2}{n-3}}(S^{n-2})}.$$

for characteristic functions f_i.

Similarly to the proof of Lemma 6.14 this boils down to proving

$$\left(\int_E\frac{r^{n-4}}{(1+r^2)^{\frac{n-2}{2}}}\,\mathrm{d}r\right)^{\frac{n-2}{n-3}} \lesssim \int_E\frac{r^{n-3}}{(1+r^2)^{\frac{n-1}{2}}}\,\mathrm{d}r$$

for all measurable $E\subseteq\mathbb{R}_+$.

To prove this we note first the following:

$$\left(\int_E\frac{1}{1+r^2}\,\mathrm{d}r\right)^{m+1} \lesssim \int_E\frac{r^m}{(1+r^2)^{\frac{m+2}{2}}}\,\mathrm{d}r.$$

To see this let $r=\tan\alpha$, then $\frac{\mathrm{d}r}{1+r^2} = \mathrm{d}\alpha$ and $(1+r^2)^{-1/2} = (1+\tan^2\alpha)^{-1/2} = \cos\alpha$ so what we want to prove is

$$\left(\int_{\widetilde{E}}\mathrm{d}\alpha\right)^{m+1} \lesssim \int_{\widetilde{E}}\tan^m\alpha\cos^m\alpha\,\mathrm{d}\alpha = \int_{\widetilde{E}}\sin^m\alpha\,\mathrm{d}\alpha.$$

In fact, we only have to prove this for $\widetilde{E} \subseteq (0, c)$ where $c > 0$ is small. In that case we can substitute the first term in its Taylor series for $\sin^m \alpha$ and then the result follows from the proof of Lemma 6.14. Now this already proves the result for $n = 4$ (take $m = 1$).

For $n > 4$ we calculate using Hölder's inequality

$$\begin{aligned}\int_{\widetilde{E}} \frac{r^{n-4}}{(1+r^2)^{\frac{n-2}{2}}}\,\mathrm{d}r &\leq \left(\int_{\widetilde{E}} \frac{r^{n-3}}{(1+r^2)^{\frac{n-1}{2}}}\,\mathrm{d}r\right)^{\frac{n-4}{n-3}} \left(\int_{\widetilde{E}} \frac{1}{1+r^2}\,\mathrm{d}r\right)^{\frac{1}{n-3}} \\ &\lesssim \left(\int_{\widetilde{E}} \frac{r^{n-3}}{(1+r^2)^{\frac{n-1}{2}}}\,\mathrm{d}r\right)^{\frac{n-4}{n-3}} \left(\int_{\widetilde{E}} \frac{r^{n-4}}{(1+r^2)^{\frac{n-2}{2}}}\,\mathrm{d}r\right)^{\frac{1}{(n-3)^2}}\end{aligned}$$

and the result follows.

Now Theorem 6.4 follows for all $n \geq 3$ by induction. □

Finally, let us return to the question how the forms are defined and prove Lemma 6.1.

Proof. To begin we take $n = 3$, the case $n = 2$ which is the Hilbert transform is of course well known. We thus want to show that

$$\frac{1}{2}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \left| \frac{(f_1(x_1) - f_1(x_1^*))f_2(x_2)f_3(x_3)}{\det\left(\begin{smallmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \end{smallmatrix}\right)} \right| \mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3$$

is bounded. We can write this as

$$\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \frac{|f_1(x_1) - f_1(x_1^*)|}{|x_1 - x_1^*|}\,\mathrm{d}x_1 \left| \frac{f_2(x_2)f_3(x_3)}{D(x_2, x_3)} \right| \mathrm{d}x_2\,\mathrm{d}x_3$$

where $D(x_2, x_3)$ is the distance between x_2 and x_3. We see that the x_1 integral is bounded as if x_1^* is close to x_1 we can estimate the integrand by $f_1'(x_1)$ and otherwise we can estimate it by a multiple of $|f_1(x_1)| + |f_1(x_1^*)|$.

For the other integrals we see that it is enough to show that

$$\int_{B_R(0)}\int_{B_R(0)} \frac{1}{|D(x_2, x_3)|}\,\mathrm{d}x_2\,\mathrm{d}x_3$$

is bounded where $B_R(0)$ denotes the ball of radius R around the origin. By letting $x_3 = x_2 + y$ we can estimate this by

$$C\int_{B_{2R}(0)} \frac{\mathrm{d}y}{|y|}$$

and by changing to polar coordinates $y = r\theta$ we can estimate this by

$$C\int_{r \leq 2R} \frac{r\,\mathrm{d}r}{r}$$

which is clearly bounded.

For the general case we proceed in the same way and we reduce our problem to showing that

$$\int_{B_R(0)} \cdots \int_{B_R(0)} \frac{1}{|D(x_2, \ldots, x_n)|} \, \mathrm{d}x_2 \ldots \mathrm{d}x_n \tag{6.22}$$

is bounded where $B_R(0)$ is a ball in $\mathbb{R}^{n-1}$ and $D(x_2, \ldots, x_n)$ is the $n-2$ dimensional volume of the simplex whose vertices are $x_2, \ldots, x_n$ in the hyperplane of $\mathbb{R}^{n-1}$ in which these points lie. As in our main argument, the boundedness of this can be shown by changing variables to separate out the contribution from x_2, changing to polar coordinates in the other variables, bounding the radial part directly and finally changing variables in the angular part to reduce to (6.22) again but with one less variable. The same argument works for Λ_S and thus we have shown the first part of the lemma.

For the second part we wish to show that

$$\int \frac{(f_1(x_1)f_2(x_2^*) - f_1(x_1^*)f_2(x_2))f_3(x_3)\cdots f_n(x_n)}{\det\left(\begin{smallmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{smallmatrix}\right)} \, \mathrm{d}x_1 \cdots \mathrm{d}x_n = 0. \tag{6.23}$$

We note that almost every tuple $(x_3, \ldots, x_n)$ lies in a uniquely determined affine plane in $\mathbb{R}^{n-1}$ of codimension 2 and we can write $x_1 = x_{10} + r(\cos(\theta)e_1 + \sin(\theta)e_2)$ and $x_2 = x_{20} + s(\cos(\phi)e_1 + \sin(\phi)e_2)$ where x_{10}, x_{20} lie in this plane and e_1 and e_2 are orthogonal unit vectors orthogonal to the plane. With these definitions we get that

$$\det\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix} = D(x_3, \ldots, x_n) rs \sin(\theta - \phi)$$

where now $D(x_3, \ldots, x_n)$ denotes the $n-3$ dimensional volume of the simplex whose vertices are $x_3, \ldots, x_n$. With this we can write the integral in (6.23) as

$$\int \frac{f(x_3)\cdots f(x_n)}{D(x_3, \ldots, x_n)} \left(\int \frac{(f_1(x_1)f_2(x_2^*) - f_1(x_1^*)f_2(x_2))}{rs\sin(\theta - \phi)} \, \mathrm{d}x_1 \, \mathrm{d}x_2 \right) \mathrm{d}x_3 \cdots \mathrm{d}x_n. \tag{6.24}$$

As above we can justify that the quantity outside of the inner integral is integrable. Let us therefore study the inner integral more carefully. We define $A_\epsilon = \{(x_1, x_2) || \sin(\theta - \phi)| > \epsilon\}$. This definition depends on the variables $x_3, \ldots, x_n$ but we shall suppress that. Note that $\lim_{\epsilon \to 0} A_\epsilon = (\mathbb{R}^{n-1})^2$ almost everywhere.

Let us study the inner integral in (6.24) restricted to the set A_ϵ. First of all note that

$$\int_{A_\epsilon} \left| \frac{f_1(x_1)f_2(x_2^*)}{rs\sin(\theta - \phi)} \right| \mathrm{d}x_1 \, \mathrm{d}x_2 \leq C \int_{\{r<R\}\cap\{s<R\}\cap A_\epsilon} \left| \frac{1}{rs\sin(\theta - \phi)} \right| rs \, \mathrm{d}r \, \mathrm{d}s \, \mathrm{d}\theta \, \mathrm{d}\phi$$

where we have carried out the x_{10} and x_{20} integrations and used the assumption that f_1 and f_2 are compactly supported. We note that the last integral is clearly bounded although the bound depends on ϵ.

For the whole inner integral restricted to A_ϵ we are therefore justified in calculating

$$\begin{aligned}\int_{A_\epsilon} &\frac{(f_1(x_1)f_2(x_2^*) - f_1(x_1^*)f_2(x_2))}{rs\sin(\theta-\phi)}\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= \int_{A_\epsilon} \frac{f_1(x_1)f_2(x_2^*)}{rs\sin(\theta-\phi)}\,\mathrm{d}x_1\,\mathrm{d}x_2 - \int_{A_\epsilon} \frac{f_1(x_1^*)f_2(x_2)}{rs\sin(\theta-\phi)}\,\mathrm{d}x_1\,\mathrm{d}x_2.\end{aligned}$$

A change of variables $x_2 \mapsto x_2^*$ in the first integral and $x_1 \mapsto x_1^*$ in the second yields

$$\int_{A_\epsilon} \frac{f_1(x_1)f_2(x_2)}{-rs\sin(\theta-\phi)}\,\mathrm{d}x_1\,\mathrm{d}x_2 - \int_{A_\epsilon} \frac{f_1(x_1)f_2(x_2)}{-rs\sin(\theta-\phi)}\,\mathrm{d}x_1\,\mathrm{d}x_2 = 0$$

Since the integral in (6.23) is absolutely integrable we get by letting ϵ pass to 0 and an application of the dominated convergence theorem that (6.24) holds. This completes the proof of the second part of the lemma and the third part is proved similarly. □

6.2 A fractional integral

Let us now look at fractional integral analogues of the form in the previous section. Define for $0 < \alpha < 1$

$$\Lambda_\alpha(f_1,\ldots,f_n) := \int \frac{f_1(x_1)\cdots f_n(x_n)}{|\det\left(\begin{smallmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n\end{smallmatrix}\right)|^\alpha}\,\mathrm{d}x_1\,\mathrm{d}x_2\ldots\,\mathrm{d}x_n \tag{6.25}$$

where $x_i \in \mathbb{R}^{n-1}$ and

$$\Lambda_{S,\alpha}(f_1,\ldots,f_n) := \int \frac{f_1(\omega_1)\cdots f_n(\omega_n)}{|\det(\omega_1,\ldots,\omega_n)|^\alpha}\,\mathrm{d}\omega_1\,\mathrm{d}\omega_2\ldots\,\mathrm{d}\omega_n \tag{6.26}$$

where $\omega_i \in S^{n-1}$. As in the Hardy–Littlewood–Sobolev theorem concerning fractional integrals, the boundedness of this multilinear form does not rely on cancellation properties of the kernel. Indeed, we have that

$$|\Lambda_\alpha(f_1,\ldots,f_n)| \lesssim \|f_1\|_{p_{0,\alpha}}\cdots\|f_{n-1}\|_{p_{0,\alpha}}\|f_n\|_1 \tag{6.27}$$

and

$$|\Lambda_{S,\alpha}(f_1,\ldots,f_n)| \lesssim \|f_1\|_{p_{0,\alpha}}\cdots\|f_{n-1}\|_{p_{0,\alpha}}\|f_n\|_1 \tag{6.28}$$

where $1/p_{0,\alpha} = 1-\alpha/(n-1)$. As before, interpolation gives that

$$|\Lambda_\alpha(f_1,\ldots,f_n)| \lesssim \|f_1\|_{p_\alpha}\cdots\|f_n\|_{p_\alpha} \tag{6.29}$$

and

$$|\Lambda_{S,\alpha}(f_1,\dots,f_n)| \lesssim \|f_1\|_{p_\alpha}\cdots\|f_n\|_{p_\alpha} \tag{6.30}$$

where $1/p_\alpha = 1-\alpha/n$. These results can be proved with the same methods we used for the singular integral version and in fact this has already been done for $\Lambda_{S,\alpha}$ by Drury [19]. Because of the absolute convergence there is no question about how the form is defined and this makes the proof slightly simpler.

There is an implied constant on the right hand side of inequalities (6.29) and (6.30). In this section we will give a minimum value for that constant and identify the functions that give equality with this constant.

The method we use is based on a treatment of the Hardy–Littlewood–Sobolev inequality in [27].

To state the theorem, let us define

$$\mathcal{H}(f_1,\dots,f_n) := \frac{|\Lambda_\alpha(f_1,\dots,f_n)|}{\|f_1\|_p\cdots\|f_n\|_p}$$

where for the rest of this section we have fixed p as p_α. Also define

$$k(x) = \frac{1}{(1+|x|^2)^{\frac{n}{2p}}} \tag{6.31}$$

We prove the following.

Theorem 6.16. *The n-tuple $(k,\dots,k)$ is an optimiser for the operator Λ_α in the sense that*

$$\sup_{f_i\geq 0}\mathcal{H}(f_1,\dots,f_n) = \mathcal{H}(k,\dots,k).$$

Furthermore, if the tuple $(f_1,\dots,f_n)$ of non-negative functions is an optimiser for Λ_α then there exists an $n\times n$ matrix A with determinant 1 and $c_i\in\mathbb{R}$ for $1\leq i\leq n$ such that

$$f_i(x) = c_i\|A\left(\begin{smallmatrix}x\\1\end{smallmatrix}\right)\|^{-\frac{n}{p}} \text{ for each } 1\leq i\leq n \tag{6.32}$$

and conversely, all tuples of functions of this form are optimisers.

Proof. Let us introduce the Steiner symmetrisation of a function. For $E\subseteq\mathbb{R}^n$ of finite Lebesgue measure we define the symmetric rearrangement of E as the open ball centred at the origin that has the same measure as E. We denote this by E^*. We then define the Steiner symmetrisation, $\mathcal{R}_j f = f^{*j}$, of a function f with respect to the j-th coordinate direction as

$$f^{*j}(x_1,\dots,x_n) = \int_0^\infty \chi_{\{|f(x_1,\dots,x_{j-1},\cdot,x_{j+1},\dots,x_n)|>t\}^*}(x_j)\,\mathrm{d}t.$$

We can see that f^{*j} is a non-negative measurable function which decreases as the absolute value of the j-th coordinate increases. Also, f and f^* have the same distribution functions and therefore $\|f\|_p = \|f^*\|_p$ for all $1 \leq p \leq \infty$. Finally, we can see that the map $f \mapsto f^*$ is order preserving, in the sense that if f and g are two non-negative functions and $f(x) \geq g(x)$ for all x then also $f^*(x) \geq g^*(x)$ for all x.

We would now like to estimate $\Lambda_\alpha(f_1, \dots, f_n)$ by $\Lambda_\alpha(f_1^*, \dots, f_n^*)$. Since

$$\det \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix}$$

is not a linear combination of the x_i's we cannot apply the rearrangement inequality of Brascamp, Lieb and Luttinger (Theorem 1.3) directly. There exists a generalisation of it by Christ [18] which is applicable (but of which we were unaware when doing this work). We shall proceed more directly.

Let us split each of the n integrals over $\mathbb{R}^{n-1}$ into integrals over $\mathbb{R}^{n-2} \times \mathbb{R}$ by separating out the integration in the j-th coordinate. Write $x_i \in \mathbb{R}^{n-1}$ as $(x_{i\hat{j}}, x_{ij})$ where x_{ij} is the j-th coordinate of x_i. Then we can write $\Lambda_\alpha(f_1, \dots, f_n)$ as

$$\int_{(\mathbb{R}^{n-2})^n} \left(\int_{\mathbb{R}^n} \frac{f_1(x_{1\hat{j}}, x_{1j}) \dots f_n(x_{n\hat{j}}, x_{nj})}{\left|\det \left(\begin{smallmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{smallmatrix} \right)\right|^\alpha} \, \mathrm{d}x_{1j} \dots \mathrm{d}x_{nj} \right) \mathrm{d}x_{1\hat{j}} \dots \mathrm{d}x_{n\hat{j}}. \tag{6.33}$$

We can work with the term in parentheses with the additional assumption that the $x_{i\hat{j}}$'s are fixed for all i's and then

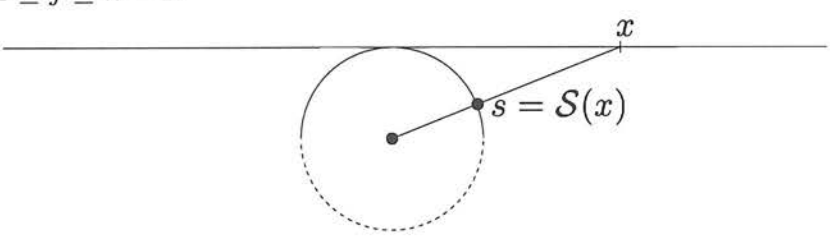
$$\det \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix}$$

is a linear combination of $x_{1j}, \dots, x_{nj}$.

We are now in a position to use Theorem 1.3. We take the functions to be $f_1(x_{1\hat{j}}, \cdot), \dots, f_n(x_{n\hat{j}}, \cdot)$, and $|\cdot|^{-\alpha}$. Now, $|\cdot|^{-\alpha}$ is a symmetric decreasing function so $(|\cdot|^{-\alpha})^* = |\cdot|^{-\alpha}$ and $f_i(x_{i\hat{j}}, \cdot)^* = f_i^{*j}(x_{i\hat{j}}, \cdot)$ where, as before, f^{*j} denotes the Steiner symmetrisation of f with respect to the j-th coordinate direction. The theorem then tells us that

$$\Lambda_\alpha(f_1, \dots, f_n) \leq \Lambda_\alpha(f_1^{*j}, \dots, f_n^{*j}) \tag{6.34}$$

for any $1 \leq j \leq n-1$.



Let $\mathcal{S} : \mathbb{R}^{n-1} \to S_+^{n-1}$ be the stereographic projection from $\mathbb{R}^{n-1}$ to the northern hemisphere S_+^{n-1}. To a function f on $\mathbb{R}^{n-1}$ we associate a function F on S_+^{n-1} defined by

$$F(s) = |J_{\mathcal{S}^{-1}}(s)|^{\frac{1}{p}} f(\mathcal{S}^{-1}(s)) \tag{6.35}$$

where $J_{\mathcal{S}^{-1}}$ is the Jacobian determinant of the map $\mathcal{S}^{-1}$. Then $\|f\|_p = \|F\|_p$ and it is easily seen that

$$\int_{(\mathbb{R}^{n-1})^n} \frac{f_1(x_1)\dots f_n(x_n)}{\left|\det\left(\begin{smallmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{smallmatrix}\right)\right|^{\alpha}} \mathrm{d}x_1 \dots \mathrm{d}x_n = \int_{(S_+^{n-1})^n} \frac{F_1(s_1)\dots F_n(s_n)}{|\det(s_1 \dots s_n)|^{\alpha}} \mathrm{d}s_1 \dots \mathrm{d}s_n. \tag{6.36}$$

We can rotate the hemisphere, by rotating the whole sphere and sending points that are rotated to the southern hemisphere to their antipodal points that lie in the northern hemisphere. The rotated functions give the same value for the integral but correspond to new functions on $\mathbb{R}^{n-1}$. We will use $U_\gamma^j f$ to denote the function we get by rotating F by the rotation that leaves all basis vectors except the j-th and the n-th ones fixed and rotates the plane spanned by those two by γ. We will require that γ is not a rational multiple of π. We note that $f \mapsto U_\gamma^j f$ is order preserving.

For a function f we define a sequence $(f^m)_{m\geq 0}$ in the following way:

$$\begin{gathered} f^0 = f, \qquad f^1 = \mathcal{R}_{n-1}\dots\mathcal{R}_1 U_\gamma^1 f^0, \\ f^2 = \mathcal{R}_1\mathcal{R}_{n-1}\dots\mathcal{R}_2 U_\gamma^2 f^1, \quad \cdots \quad f^{n-1} = \mathcal{R}_{n-2}\dots\mathcal{R}_1\mathcal{R}_{n-1} U_\gamma^{n-1} f^{n-2}, \\ f^n = \mathcal{R}_{n-1}\dots\mathcal{R}_1 U_\gamma^1 f^{n-1}, \quad \cdots . \end{gathered}$$

We want to find the L^p limit of this sequence. First, let us assume that f is a bounded function which vanishes outside a bounded set. These functions are clearly dense in L^p. With this assumption we can find a constant C such that

$$f(x) \leq C k_f(x) \tag{6.37}$$

where $k_f(x)$ is a multiple of $k(x)$ from (6.31) scaled such that $\|f\|_p = \|k_f\|_p$. We notice that $k_f(x)$ is a symmetric decreasing function which corresponds to a constant function K on on S_+^{n-1}. It is thus unaffected by $\mathcal{R}_j$ and U_γ^j. Since $f(x) > 0$ and both $\mathcal{R}_j$ and U_γ^j preserve orderings of non-negative functions we have that

$$f^m(x) \leq C k_f^m(x) = C k_f(x) \tag{6.38}$$

for all x and m so the whole sequence (f^m) is dominated by an L^p function. Since

$$\|k_f - U_\gamma^j f\|_p = \|U_\gamma^j k_f - U_\gamma^j f\|_p = \|k_f - f\|_p \tag{6.39}$$

and

$$\|k_f - \mathcal{R}_j f\|_p = \|\mathcal{R}_j k_f - \mathcal{R}_j f\|_p \leq \|k_f - f\|_p \tag{6.40}$$

by Theorem 1.5 we have that

$$\lim_{m\to\infty} \|k_f - f^m\| \tag{6.41}$$

exists and is equal to

$$\inf_m \|k_f - f^m\|. \tag{6.42}$$

We call this number A. It is finite since $\|k_f - f\|_p \leq \|k_f\|_p + \|f\|_p < \infty$.

We make the following definition:

Definition 6.17. Let f be a non-negative function. We say that f has the *outward decreasing property* if for all $x, y \in \mathbb{R}^{n-1}$ s.t. $|x_i| \leq |y_i|$ for all $1 \leq i \leq n$ then $f(x) \geq f(y)$.

Lemma 6.18. *The functions f^m have the outward decreasing property for $m \geq 1$.*

Proof. It is clear that we can take $x_i, y_i \geq 0$ for all i and that it is enough to show that $g = \mathcal{R}_1\mathcal{R}_2 f$ has the property that if $x_1 \leq y_1$, $x_2 \leq y_2$ and $x_i = y_i$ for $i \geq 3$ then $g(x) \geq g(y)$.

Furthermore, since $g = \mathcal{R}_1(\mathcal{R}_2 f)$ it is clear that if we also have $x_2 = y_2$ then $g(x) \geq g(y)$. So it is enough to study the case $x_1 = y_1$, $x_i = y_i$ for $i \geq 3$ and $x_2 \leq y_2$. Obviously, in this case,

$$\mathcal{R}_2 f(x) \geq \mathcal{R}_2 f(y). \tag{6.43}$$

Now set $\lambda := g(y)$. Then

$$|\{t : g(t, y_2, \dots, y_{n-1}) \geq \lambda\} = 2y_1$$

so

$$|\{t : \mathcal{R}_2 f(t, y_2, \dots, y_{n-1}) \geq \lambda\} = 2y_1.$$

Since $x_2 \leq y_2$ we have that

$$\mathcal{R}_2 f(t, x_2, y_3, \dots, y_{n-1}) \geq \mathcal{R}_2 f(t, y_2, y_3, \dots, y_{n-1})$$

for all $t, y_3, \dots, y_{n-1}$ so

$$|\{t : \mathcal{R}_2 f(t, x_2, y_3, \dots, y_{n-1}) \geq \lambda\} \geq 2y_1$$

which is

$$|\{t : \mathcal{R}_2 f(t, x_2, x_3, \dots, x_{n-1}) \geq \lambda\} \geq 2x_1$$

and this tells us that

$$g(x) = \mathcal{R}_1\mathcal{R}_2 f(x_1, \dots, x_{n-1}) \geq \lambda$$

so $g(x) \geq g(y)$. This completes the proof of the lemma. $\square$

Using this property and Helly's selection principle we can find a subsequence f^{m_j} which converges to some h almost everywhere. We can also impose the condition that $(n-1)|m_j$ for all j. It is clear that this h will also have the outward decreasing property. Since all the functions f^m are dominated by the L^p function Ch_f we see that h belongs to L^p and

$$A = \lim_{j\to\infty} \|f^{m_j} - k_f\|_p = \|h - k_f\|_p. \tag{6.44}$$

However, we also have

$$A = \lim_{j\to\infty} \|f^{m_j+1} - k_f\|_p = \|\mathcal{R}_{n-1}\dots\mathcal{R}_1 U_\gamma^1 h - k_f\|_p$$

so

$$\begin{aligned} A &\le \|\mathcal{R}_{n-1}\dots\mathcal{R}_1 U_\gamma^1 h - k_f\|_p = \|\mathcal{R}_{n-1}\dots\mathcal{R}_1 U_\gamma^1 h - \mathcal{R}_{n-1}\dots\mathcal{R}_1 U_\gamma^1 k_f\|_p \\ &\le \|U_\gamma^1 h - U_\gamma^1 k_f\|_p = \|h - k_f\|_p = A \end{aligned}$$

which tells us that we must have equality everywhere in the chain. In particular,

$$\|\mathcal{R}_1 U_\gamma^1 h - k_f\|_p = \|U_\gamma^1 h - k_f\|_p.$$

Now Theorem 1.5 tells us that for almost every $x_2, \dots, x_{n-1}$ we have that

$$\mathcal{R}_1 U_\gamma^1 h(x) = U_\gamma^1 h(x).$$

Thus we have shown that both h and $U_\gamma h$ are invariant under the reflection

$$h(x_1, x_2, \dots, x_{n-1}) \mapsto \mathcal{T}^1 h := h(-x_1, x_2, \dots, x_{n-1})$$

and since $U_{-\gamma}^1 h = \mathcal{T}^1 U_\gamma \mathcal{T}^1 h$ we see that $U_{-\gamma}^1 h = U_\gamma^1 h$ so $U_{2\gamma}^1 h = U_\gamma^1 U_\gamma^1 h = U_\gamma^1 U_{-\gamma}^1 h = h$. Since γ is not a rational multiple of π we see that H, the function on the northern hemisphere associated to h, is constant along curves which are intersections of the northern hemisphere and translates of the $x_1 x_n$ coordinate plane. This also tells us that

$$h = U_\gamma^1 h = \mathcal{R}_1 U_\gamma^1 h \quad \text{a.e.} \tag{6.45}$$

Now we can use the chain of equalities

$$\|\mathcal{R}_{n-1}\dots\mathcal{R}_1 U_\gamma^1 h - k_f\|_p = \dots = \|\mathcal{R}_1 U_\gamma^1 h - k_f\|_p = \|U_\gamma^1 h - k_f\|_p \tag{6.46}$$

to see that

$$\mathcal{R}_{n-1}\dots\mathcal{R}_1 U_\gamma^1 h = \dots = \mathcal{R}_1 U_\gamma^1 h = U_\gamma^1 h = h \quad \text{a.e.} \tag{6.47}$$

We also have

$$\mathcal{R}_1 U_\gamma^2 \mathcal{R}_{n-1} \dots \mathcal{R}_1 U_\gamma^1 h = U_\gamma^2 \mathcal{R}_{n-1} \dots \mathcal{R}_1 U_\gamma^1 h \tag{6.48}$$

so the same argument tells us that the function on the northern hemisphere associated to $\mathcal{R}_{n-1} \dots \mathcal{R}_1 U_\gamma^1 h$ is constant along curves which are intersections of the northern hemisphere and translates of the $x_2 x_n$ coordinate plane. Since $\mathcal{R}_{n-1} \dots \mathcal{R}_1 U_\gamma^1 h = h$ a.e. we see that H is a.e. constant on 3-spaces which are parallel to the $x_1 x_2 x_n$-coordinate 3-space.

From this discussion the induction is evident and the result will be that H is a.e. constant on the northern hemisphere and since h has the outward decreasing property we see that H must be constant everywhere and h must have the form Ck_f for some C. Since $\|h\|_p = \lim_{j\to\infty} \|f^{m_j}\|_p = \|f\|_p = \|k_f\|_p$ we see that $C = 1$ and $h = k_f$.

This tells us that $A = 0$ and since $(\|k_f - f^m\|_p)_{m=0}^\infty$ is a decreasing sequence with a subsequence which tends to 0 we see that the whole sequence (f^m) tends to k_f. We have thus shown that for any f in the dense class of L^p functions we started with that $f^m \to k_f$. Since $\|k_f - k'_f\|_p \le \|f - f'\|_p$ for any $f, f' \in L^p$ we see that for any $f \in L^p$ we have that $f^m \to k_f$ in L^p.

Now

$$\mathcal{H}(f_1^m, \dots, f_n^m) \le \mathcal{H}(f_1^{m+1}, \dots, f_n^{m+1}) \tag{6.49}$$

for every $m \ge 0$ so

$$\mathcal{H}(f_1, \dots, f_n) \le \mathcal{H}(k_f, \dots, k_f) = \mathcal{H}(k, \dots, k). \tag{6.50}$$

This tells us that $(k, \dots, k)$ is an optimiser for Λ_α.

Now let us find all the non-negative functions which furnish the best constant. Using Lemma 1.4 we can see that

$$\Lambda_\alpha(f_1, \dots, f_n) = \Lambda_\alpha(f_1^{*j}, \dots, f_n^{*j})$$

can hold only if $f_i(x) = f_i^*(x - a_j e_j)$ where e_j is the j-th coordinate vector and the a_j's satisfy

$$\det \begin{pmatrix} 1 & \dots & 1 \\ x_{11} & \dots & x_{n1} \\ \vdots & & \vdots \\ x_{1,j-1} & \dots & x_{n,j-1} \\ a_1 & \dots & a_n \\ x_{1,j+1} & \dots & x_{n,j+1} \\ \vdots & & \vdots \\ x_{1,n-1} & \dots & x_{n,n-1} \end{pmatrix} = 0.$$

This conclusion holds provided that all the adjoint matrices of the a_i's are nonzero and that is true for almost any $x_1, \dots, x_n \in \mathbb{R}^{n-1}$.

Now, let us say that for some $x_{2\hat{j}}, \dots, x_{n\hat{j}}$, where we do not specify the j-th coordinate in each vector, we have found that $f_i(x_{i\hat{j}}, \cdot)$ has centre at a_i for $2 \le i \le n$. Then we can see that for any $x_{i\hat{j}}$ the centre of $f_1(x_{1\hat{j}}, \cdot)$ must be at the point a_1 such that all the $(x_{i\hat{j}}, a_i)$ lie in some $(n-2)$-dimensional hyperplane. Then, by moving the $x_{i\hat{j}}$'s around one by one for $2 \le i \le n$ we can see that there must exist a hyperplane where all the points $(x_{i\hat{j}}, a_i)$ lie.

This tells us that if $(f_1, \dots, f_n)$ is an optimiser for our operator then the functions have the form $f_i(x_i) = h_i(Mx_i + b)$ where the h_i's have the outward decreasing property, M is an $(n-1) \times (n-1)$ matrix A with determinant 1 and $b \in \mathbb{R}^{n-1}$.

Now, the transformations $f \mapsto U^j_\alpha f$ and $f \mapsto f(M \cdot + b)$ span a group G. It is now clear that for an optimiser $(f_1, \dots, f_n)$ the rearrangements $\mathcal{R}_j f_i$ will be of the form $T_g f$ for some $g \in G$ and thus the whole sequence $(f^m)_{m \ge 0}$ will be of the form $T_{g_m} f_i$ for some $g_m \in G$, the same g_m for each i.

Since the elements of G are isometries of L^p we have that

$$0 = \lim_{k \to \infty} \|f_i^m - k_{f_i}\|_p = \lim_{k \to \infty} \|f_i - T_{g_m^{-1}} k_{f_i}\|_p.$$

We shall see that for any $g \in G$ we have

$$T_g k_f(x) = \left(\begin{pmatrix} x \\ 1 \end{pmatrix}^T A^T A \begin{pmatrix} x \\ 1 \end{pmatrix} \right)^{-\frac{n}{2p}} \tag{6.51}$$

for some real $n \times n$ matrix with determinant 1.

Let $f_A(x) = \|A \left(\begin{smallmatrix} x \\ 1 \end{smallmatrix}\right)\|^{-\frac{n}{p}}$. Then $f_A(Mx + b) = \|A \left(\begin{smallmatrix} Mx+b \\ 1 \end{smallmatrix}\right)\|^{-\frac{n}{p}} = \|A' \left(\begin{smallmatrix} x \\ 1 \end{smallmatrix}\right)\|^{-\frac{n}{p}}$ with $A' = A \left(\begin{smallmatrix} M & b \\ 0 & 1 \end{smallmatrix}\right)$ so A' is again a real $n \times n$ matrix with determinant 1.

Now consider U^j_α for some j. Without loss of generality we can take $j = 1$. Then

$$U^1_\alpha f_A(x) = \left(\frac{1}{1 + |x|^2} \frac{1 + |w|^2}{\|A \left(\begin{smallmatrix} w \\ 1 \end{smallmatrix}\right)\|^2} \right)^{\frac{n}{2p}}$$

where $\left(\begin{smallmatrix} w \\ 1 \end{smallmatrix}\right)$ is the point in $\mathbb{R}^n$ we get by starting with $\left(\begin{smallmatrix} x \\ 1 \end{smallmatrix}\right)$, projecting it to the hemisphere, that is, to $\frac{1}{\sqrt{1+|x|^2}} \left(\begin{smallmatrix} x \\ 1 \end{smallmatrix}\right)$, then rotating in the $x_1 x_n$-plane by $-\alpha$, this sends

$$\frac{1}{\sqrt{1 + |x|^2}} \begin{pmatrix} x \\ 1 \end{pmatrix} = \frac{1}{\sqrt{1 + |x|^2}} \begin{pmatrix} x_1 \\ \hat{x}_1 \\ 1 \end{pmatrix}$$

to

$$\frac{1}{\sqrt{1 + |x|^2}} \begin{pmatrix} \cos \alpha x_1 + \sin \alpha \\ \hat{x}_1 \\ -\sin \alpha x_1 + \cos \alpha \end{pmatrix},$$

and finally projecting this point to the plane $\left(\begin{smallmatrix} w \\ 1 \end{smallmatrix}\right)$, which sends it to

$$\begin{pmatrix} (\cos\alpha x_1 + \sin\alpha)/(-\sin\alpha x_1 + \cos\alpha) \\ \hat{x}_1/(-\sin\alpha x_1 + \cos\alpha) \\ 1 \end{pmatrix};$$

so $w = (-\sin\alpha x_1 + \cos\alpha)^{-1}\left(\begin{smallmatrix} \cos\alpha x_1 + \sin\alpha \\ \hat{x}_1 \end{smallmatrix}\right) =: w_n^{-1}\left(\begin{smallmatrix} w_1 \\ \hat{x}_1 \end{smallmatrix}\right)$. Since $w_1^2 + w_n^2 = x_1^2 + 1$ we have that

$$\left(\frac{1}{1+|x|^2}\frac{1+|w|^2}{\|A\left(\begin{smallmatrix} w \\ 1 \end{smallmatrix}\right)\|^2}\right)^{\frac{n}{2p}} = \left(\frac{1}{1+|x|^2}\frac{1+x_1^2+|\hat{x}_1|^2}{\left\|A\left(\begin{smallmatrix} w_1 \\ \hat{x}_1 \\ w_n \end{smallmatrix}\right)\right\|^2}\right)^{\frac{n}{2p}} = \left\|A\left(\begin{smallmatrix} w_1 \\ \hat{x}_1 \\ w_n \end{smallmatrix}\right)\right\|^{-\frac{n}{p}}$$

and since

$$\begin{pmatrix} w_1 \\ \hat{x}_1 \\ w_n \end{pmatrix} = \begin{pmatrix} \cos\alpha x_1 + \sin\alpha \\ \hat{x}_1 \\ -\sin\alpha x_1 + \cos\alpha \end{pmatrix} = \begin{pmatrix} \cos\alpha & 0 & \sin\alpha \\ 0 & I & 0 \\ -\sin\alpha & 0 & \cos\alpha \end{pmatrix}\begin{pmatrix} x_1 \\ \hat{x}_1 \\ 1 \end{pmatrix}$$

we get

$$U_\alpha^1 f_A(x) = \frac{1}{\|A'\left(\begin{smallmatrix} x \\ 1 \end{smallmatrix}\right)\|^{\frac{n}{p}}}$$

with

$$A' = A\begin{pmatrix} \cos\alpha & 0 & \sin\alpha \\ 0 & I & 0 \\ -\sin\alpha & 0 & \cos\alpha \end{pmatrix}$$

and again A' has determinant 1.

Since the set of functions $\{\|A\left(\begin{smallmatrix} x \\ 1 \end{smallmatrix}\right)\|^{-\frac{n}{p}} | A \text{ is a } n\times n \text{ matrix with } \det A = 1\}$ is closed in L^p and k_f belongs to this set we have shown that all optimisers have the form prescribed in the theorem.

Let us now see that all functions of the prescribed form are optimisers. It is clear that we can take $c_i = 1$. Let us therefore again take $f_A := \|A\left(\begin{smallmatrix} x \\ 1 \end{smallmatrix}\right)\|^{-\frac{n}{p}}$. Then it is enough to show that $\|f_A\|_p = \|f_I\|_p$ and

$$\Lambda_\alpha(f_A, \ldots, f_A) = \Lambda_\alpha(f_I, \ldots, f_I)$$

where I is the $n\times n$ identity matrix because we know that $(f_I)_{i=1}^n$ is an optimiser. To prove the equality we note first of all that $\Lambda_\alpha(f_1, \ldots, f_n)$ is invariant under the transformation $(f_1, \ldots, f_n) \mapsto (f_1(M\cdot + b), \ldots, f_n(M\cdot + b))$ where as before M is an $(n-1)\times(n-1)$ matrix with determinant 1 and $b \in \mathbb{R}^{n-1}$. We note also that these transformations preserve the L^p-norm of the functions. By using this invariance we may make the additional assumption that A has the form $\left(\begin{smallmatrix} d_1 & 0 \\ 0 & d_2^2 I \end{smallmatrix}\right)$ with positive scalars d_1 and d_2 where I denotes the identity matrix of size $n-1$. Since we have that $\det A = 1$ we get the relation $d_1(d_2)^{2(n-1)} = \det A = 1$.

So we want to consider $\Lambda_\alpha(f_A, \dots, f_A)$ which equals

$$\int \frac{\left(\left(\frac{1}{d_2^{2(n-1)}} + \|d_2 x_1\|^2\right) \dots \left(\frac{1}{d_2^{2(n-1)}} + \|d_2 x_n\|^2\right)\right)^{-\frac{n}{2p}}}{\left|\left(\begin{smallmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{smallmatrix}\right)\right|^\alpha} \, \mathrm{d}x_1 \dots \mathrm{d}x_n.$$

We make the change of variables $d_2^n x_i = y_i$. Then $d_2^{n(n-1)} \, \mathrm{d}x_i = \mathrm{d}y_i$ and

$$\det \begin{pmatrix} 1 & \dots & 1 \\ y_1 & \dots & y_n \end{pmatrix} = d_2^{n(n-1)} \det \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix}.$$

We thus get

$$\begin{aligned} \Lambda_\alpha(f_A, \dots, f_A) &= \int \frac{\left(\frac{1}{d_2^{2n(n-1)}}\right)^{-\frac{n}{2p}} ((1 + \|y_1\|^2) \dots (1 + \|y_n\|^2))^{-\frac{n}{2p}}}{d_2^{-n(n-1)\alpha} \left|\det \left(\begin{smallmatrix} 1 & \cdots & 1 \\ y_1 & \cdots & y_n \end{smallmatrix}\right)\right|^\alpha} \frac{\mathrm{d}y_1 \dots \mathrm{d}y_n}{d_2^{n^2(n-1)}} \\ &= d_2^{(n-1)\left(\frac{n^2}{p} - n^2 + n\alpha\right)} \Lambda_\alpha(f_I, \dots, f_I). \end{aligned}$$

Now note that $\frac{n^2}{p} - n^2 + n\alpha = n^2(1 - \frac{\alpha}{n}) - n^2 + n\alpha = 0$ so that we have the desired equality of the forms.

Finally, we calculate

$$\begin{aligned} \|f_A\|_p &= \int_{\mathbb{R}^{n-1}} \left(\frac{1}{d_2^{2(n-1)}} + \|d_2 x\|^2\right)^{-\frac{n}{2}} \mathrm{d}x \\ &= \int_{\mathbb{R}^{n-1}} \left(1 + \|d_2^n x\|^2\right)^{-\frac{n}{2}} d_2^{n(n-1)} \, \mathrm{d}x = \|f_I\|_p \end{aligned}$$

where we have used the same change of variables as above. This completes the proof. □

Let us now examine the form $\Lambda_{S,\alpha}$ defined in (6.26). We have for any functions $\tilde{F}_i$ defined on S^{n-1} that

$$\Lambda_{S,\alpha}(\tilde{F}_1, \dots, \tilde{F}_n) = \int_{(S_+^{n-1})^n} \frac{F_1(s_1) \dots F_n(s_n)}{|\det(s_1 \dots s_n)|^\alpha} \, \mathrm{d}s_1 \dots \mathrm{d}s_n =: \tilde{\Lambda}_{S,\alpha}(F_1, \dots, F_n)$$

where $F_i(s_i) = \tilde{F}_i(s_i) + \tilde{F}_i(\bar{s}_i)$ and $\bar{s}_i$ is the antipodal point of s_i.

We note from this proof that $(f_1, \dots, f_n)$ is an optimiser for Λ_α if and only if $(F_1, \dots, F_n)$ is an optimiser for $\tilde{\Lambda}_{S,\alpha}$ where f_i and F_i are related by (6.35). Furthermore, for any $s = (s_1, \dots, s_n) \in S_+^{n-1}$ we have that $\mathcal{S}^{-1}(s) = s/s_n$ and $|J_{\mathcal{S}^{-1}}(s)| = s_n^{-n}$ by the same calculation as in the previous section so if f_i has the form $f_i(x) = c_i \|A\left(\begin{smallmatrix} x \\ 1 \end{smallmatrix}\right)\|^{-\frac{n}{p}}$ as in (6.32) then the corresponding F_i has the form $F_i(s) = c_i \|As\|^{-\frac{n}{p}}$. Thus we can state the analogue of Theorem 6.16 for $\Lambda_{S,\alpha}$ as follows.

Theorem 6.19. *The n-tuple of constant functions is an optimiser for the operator $\Lambda_{S,\alpha}$ and if the tuple $(\tilde{F}_1, \ldots, \tilde{F}_n)$ of non-negative functions is an optimiser for $\Lambda_{S,\alpha}$ then there exists an $n \times n$ matrix A with determinant 1 and $c_i \in \mathbb{R}$ for $1 \leq i \leq n$ such that*

$$F_i(s) = c_i \|As\|^{-\frac{n}{p}} \text{ for each } 1 \leq i \leq n \tag{6.52}$$

where $F_i : S_+^{n-1} \to \mathbb{R}$ are given by $F_i(s) = \tilde{F}_i(s) + \tilde{F}_i(\bar{s})$ and $\bar{s}$ is the antipodal point of s. Conversely, all tuples of functions satisfying this condition are optimisers.

Bibliography

[1] Keith Ball. Volumes of sections of cubes and related problems. In *Geometric aspects of functional analysis (1987–88)*, volume 1376 of *Lecture Notes in Math.*, pages 251–260. Springer, Berlin, 1989.

[2] Keith Ball. An elementary introduction to monotone transportation. In *Geometric aspects of functional analysis*, volume 1850 of *Lecture Notes in Math.*, pages 41–52. Springer, Berlin, 2004.

[3] Franck Barthe. On a reverse form of the Brascamp–Lieb inequality. *Invent. Math.*, 134(2):335–361, 1998.

[4] Alexander Barvinok. *A course in convexity*, volume 54 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002.

[5] William Beckner. Inequalities in Fourier analysis. *Ann. of Math. (2)*, 102(1):159–182, 1975.

[6] Jonathan Bennett, Anthony Carbery, Michael Christ, and Terence Tao. Finite bounds for Hölder–Brascamp–Lieb multilinear inequalities. *Math. Res. Lett.* to appear.

[7] Jonathan Bennett, Anthony Carbery, Michael Christ, and Terence Tao. The Brascamp–Lieb inequalities: finiteness, structure, and extremals. *Geom. Funct. Anal.* to appear.

[8] Herm J. Brascamp and Elliott H. Lieb. Best constants in Young's inequality, its converse, and its generalization to more than three functions. *Advances in Math.*, 20(2):151–173, 1976.

[9] Herm J. Brascamp, Elliott H. Lieb, and Joaquin M. Luttinger. A general rearrangement inequality for multiple integrals. *J. Functional Analysis*, 17:227–237, 1974.

[10] Yann Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I Math.*, 305(19):805–808, 1987.

[11] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

[12] Almut Burchard. *Cases of equality in the Riesz rearrangement inequality.* PhD thesis, School of Mathematics, Georgia Institute of Technology, May 1994.

[13] Almut Burchard. Cases of equality in the Riesz rearrangement inequality. *Ann. of Math. (2)*, 143(3):499–527, 1996.

[14] Luis A. Caffarelli. Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.*, 214(3):547–563, 2000.

[15] Luis A. Caffarelli. Erratum: "Monotonicity of optimal transportation and the FKG and related inequalities" [Comm. Math. Phys. **214** (2000), no. 3, 547–563; MR1800860 (2002c:60029)]. *Comm. Math. Phys.*, 225(2):449–450, 2002.

[16] Eric A. Carlen, Elliott H. Lieb, and Michael Loss. A sharp analog of Young's inequality on S^N and related entropy inequalities. *J. Geom. Anal.*, 14(3):487–520, 2004.

[17] Eric A. Carlen and Michael Loss. Extremals of functionals with competing symmetries. *J. Funct. Anal.*, 88(2):437–456, 1990.

[18] Michael Christ. Estimates for the k-plane transform. *Indiana Univ. Math. J.*, 33(6):891–910, 1984.

[19] Stephen W. Drury. Estimates for a multilinear form on the sphere. *Math. Proc. Cambridge Philos. Soc.*, 104(3):533–537, 1988.

[20] Wilfrid Gangbo and Robert J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.

[21] Loukas Grafakos and Nigel Kalton. Some remarks on multilinear maps and interpolation. *Math. Ann.*, 319(1):151–180, 2001.

[22] Roger A. Horn and Charles R. Johnson. *Matrix analysis.* Cambridge University Press, Cambridge, 1985.

[23] Martin Knott and Cyril S. Smith. On the optimal mapping of distributions. *J. Optim. Theory Appl.*, 43(1):39–49, 1984.

[24] Elliott H. Lieb. Existence and uniqueness of the minimizing solution of Choquard's nonlinear equation. *Studies in Appl. Math.*, 57(2):93–105, 1976/77.

[25] Elliott H. Lieb. Sharp constants in the Hardy-Littlewood-Sobolev and related inequalities. *Ann. of Math. (2)*, 118(2):349–374, 1983.

[26] Elliott H. Lieb. Gaussian kernels have only Gaussian maximizers. *Invent. Math.*, 102(1):179–208, 1990.

[27] Elliott H. Lieb and Michael Loss. *Analysis*, volume 14 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1997.

[28] Robert J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80(2):309–323, 1995.

[29] Gaspard Monge. Mémoire sure la théorie des déblais et des remblais. In *Histoire de l'Académie Royale des Sciences de Paris*, pages 666–704. 1781.

[30] Gian-Carlo Rota. The many lives of lattice theory. *Notices Amer. Math. Soc.*, 44(11):1440–1445, 1997.

[31] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.